32060

National Library of Canada
Bibliothèque nationale du Canada

CANADIAN THESES ON MICROFICHE
THÈSES CANADIENNES SUR MICROFICHE

NAME OF AUTHOR/*NOM DE L'AUTEUR* Geoffrey Tobin Rowe

TITLE OF THESIS/*TITRE DE LA THÈSE* Ambiguity And Uncertainty: An Experimental Study

UNIVERSITY/*UNIVERSITÉ* University of Alberta

DEGREE FOR WHICH THESIS WAS PRESENTED/ *GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE* Master of Science

YEAR THIS DEGREE CONFERRED/*ANNÉE D'OBTENTION DE CE GRADE* 1977

NAME OF SUPERVISOR/*NOM DU DIRECTEUR DE THÈSE* Dr. Gary D. Prideaux

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

DATED/*DATÉ* April 22, 1977 SIGNED/*SIGNÉ* Geoff Rowe

PERMANENT ADDRESS/*RÉSIDENCE FIXÉ* 10320 - 124 St.
Edmonton, Alberta
Canada.

NL-91 (3-74)

National Library of Canada

Cataloguing Branch
Canadian Theses Division

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

Bibliothèque nationale du Canada

Direction du catalogage
Division des thèses canadiennes

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

NL-339 (3/77)

THE UNIVERSITY OF ALBERTA

AMBIGUITY AND UNCERTAINTY: An Experimental Study

by

© GEOFFREY TOBIN ROWE

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE
OF MASTER OF SCIENCE
IN
PSYCHOLINGUISTICS

DEPARTMENT OF LINGUISTICS

EDMONTON, ALBERTA
SPRING, 1977

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled AMBIGUITY AND UNCERTAINTY: AN EXPERIMENTAL STUDY

submitted by Geoffrey Tobin Rowe

in partial fulfilment of the requirements for the degree of Master of Science.

Gary D Prideaux
Supervisor

Lois Marckworth

Anton Rozsypal

John T. Hogan

William T. Fagan
External Examiner

Date March 14, 1977

ABSTRACT

The relative difficulty (as represented by variations in response times) that speakers have in identifying the alternative interpretations of ambiguous sentences has been represented as an instance of grammatically rule governed sentence comprehension processes. Prominent among such research is the MacKay - Bever (1967) study. MacKay and Bever's paper purports to establish:

i. a direct relation between grammatical sentence properties, as characterized by transformational grammar, and comprehension response times,

and

ii. the empirical status of a taxonomy of ambiguity, for which the theortical distinction between types depends on a process interpretation of transformational grammar.

In respect to the latter claim, it was noted that it assumed that the hypothetical distinction between the grammatical constructs, 'Surface Structure' and 'Deep Structure', have special psychological as well as formal status. Thus, it was concluded that MacKay and Bever's argument was circular and their interpretation of their results was in jeopardy.

This thesis was intended to replicate and extend the MacKay - Bever study. Their procedures were modified to take into account several perceived weaknesses, among which were: (i) the absence of nonambiguous control sentences in the stimulus set, (ii) the absence of control for the semantic

and syntactic heterogeneity of the stimulus items, and (iii) the inadequacy of their statistical analysis.

The experimental procedure involved presentation of the stimuli to subjects under a computer managed regime. The presentation program included automatic randomization of stimuli and automatic recording of the response data. The stimuli were adapted from MacKay and Bever's stimulus list. However, each ambiguous stimulus was modified to contain a set number of words and to have two distinct clauses. For each ambiguous stimulus, a corresponding nonambiguous sentence was constructed having an analogous syntactic structure. Six stimulus lists were compiled. In each list there were equal numbers of instances of the stimulus types: Lexically Ambiguous, Surface Structure Ambiguity, Underlying Ambiguity, and Nonambiguous. Three of the lists contained stimuli with their clauses in one order, while the other three lists contained the same stimuli but with their clauses in the reverse order.

The results of the experiment reported in this thesis have not born out the MacKay - Bever interpretations of the sentence comprehension process. The follwing discrepancies were identified:

i. The major source of variation was intersubjective,

ii. no single profile of response times across stimulus types was charateristic of the subjects' responses, and thus no unique response strategy could be inferred.

iii. no empirical basis was found for distinguishing, in terms of response times, between the types of structural ambiguity identified by MacKay and Bever with 'Surface' and 'Deep' structure,

iv. evidence was found of a relation between speed and accuracy of response that is characteristic of statistical decision-making proceses rather than of deterministic rule-governed processes.

An examination of the experimental paradigm and of the theoretical basis for this type of research led to the implication that the behaviour was related to the subject matter of Choice Reaction Time Theory. The importance of so associating the research with established psychological theory lay in the recognition that the behaviour need not be construed as an exclusively language-related phenomenon, independent of more general human information processing abilities.

It was concluded that as with so many other information processing tasks, the problem of interpreting a sentence's meaning involves decision under uncertainty. This effect appears to be sufficiently general to suggest that from the point of view of a hearer every sentence must be regarded as potentially ambiguous.

Acknowledgements

This work would have been impossible without the contributions and long-suffering patience of my supervisor, Dr. Gary D. Prideaux, and the other members of my thesis committee. I would particularly like to thank Dr. John T. Hogan for his invaluable assistance in data analysis, and to thank Dr. Anton Rozsypal who provided me with the stimulus presentation program. Dr. Lois Marckworth deserves special credit for helping to guide me through the stylistic quagmire of the preliminary drafts. I would also like to thank the remaining member of my committee, Dr. William T. Fagan, who provided insightful criticism.

My study towards the masters degree was supported by the University of Alberta and the Department of Linguistics through graduate assistantships and intersession bursaries.

TABLE OF CONTENTS

CHAPTER PAGE

I. Chapter One - Introduction

1. Preliminaries 1

An Overview of the Issues 4

Grammatical Analysis and Message Processing 6

Meaning and 'Normal' Processing 7

2. Transformational Grammars and the Classification of Ambiguity 11

Justification of Transformational Grammar 11

Classification and Judgement of Ambiguity 14

Problems for the Interpretation of Experimental Results....................................17

3. The MacKay-Bever Experiment 18

Purposes and Procedure 18

Results 19

4. A Critique of the MacKay-Bever Experiment 22

Analysis and Design 22

Methodology 26

Interpretation 27

The Character of the Dependent Variable 28

5. The Relation between Stimulus Categorization and Latency 31

The Choice Reaction Time Paradigm 31

CRT Processing Models 34

Ambiguity and the CRT Paradigm 37

Psycholinguistic Research on Ambiguity 40

TABLE OF CONTENTS CONTINUED

II. Chapter Two - Method 49

1. Experiment 1 50

  Subjects 50

  Apparatus 51

  Materials 52

  Procedure 54

  Design 55

2. Experiment 2 57

  Subjects 57

  Materials 57

  Procedure 58

III. Chapter Three - Results

1. Preliminaries 59

2. Response Time Analysis 61

  Multivariate Analysis of Variance 62

  Discriminant Function Analysis 66

2. Analysis of Covariates 73

  The Covariates 72

  Multiple Regression 77

IV. Chapter Four - Discussion

1. Sentence Processing 81

  Ambiguity Type and Stages of Processing 81

  Order of Processing 84

  Processing: Deterministic or Probabilistic 87

2. Relation to Issues in the Literature 94

3. Difficulty and CPT Theory 96

4. Processing Dynamics and Linguistic Theory 98

TABLE OF CONTENTS CONTINUED

V. Chapter Five - Summary and Conclusion 101

References 107

Appendix A - Experimental Instructions 112

Appendix B - Stimulus Set 114

Appendix C - Statistical Results 117

LIST OF FIGURES

FIGURE PAGE
1. Data Structure 121
2. Discriminators Response Profile 122
3. Group 1 Response Profile 123
4. Group 2 Response Profile 124
5. Group 3 Response Profile 125
6. Group 4 Response Profile 126
7. Response Time Frequency Distribution 127

CHAPTER ONE

INTRODUCTION

Preliminaries

The purpose of this thesis is to replicate and extend an experiment (MacKay and Bever, 1967) which has played a central role in the psycholinguistic research dealing with sentence comprehension. This experiment involved observations of response times leading to comprehension of the alternative meanings of ambiguous sentences. Under the assumption that linguistically defined properties of the sentences contributed directly to the variations among the response times, this experiment was interpreted as revealing the grammatical basis of the sentence comprehension process. The results were identified as being consistent with the recognition of three uniquely ordered stages of processing, each stage corresponding to a level of analysis in a transformational derivation of the sentences. The stages were associated with categories of ambiguity which, putatively, could only be resolved at that level of analysis. Thus, the stages of processing might be identified as lexical, surface structure, and deep structure information processing.

By way of introduction, five areas of potential controversy will be discussed. These are:

1) The place of formally defined sentence properties

among the diverse influences on the behaviour of naive speakers.

2) The treatment of ambiguity in transformational theory and the formal status of the MacKay-Bever taxonomy of ambiguities.

3) The conceptual, methodological, and analytical weaknesses of the MacKay-Bever experiment.

4) The considerations involved in the use of reaction time as a dependent measure in tasks requiring stimulus categorization.

5) The difficulties to be faced in distinguishing psycholinguistically interesting effects from theoretically irrelevant sources of variation.

Among the conclusions reached is the realization that differentiation among proposed formal properties of sentences alone may not serve as an explanation of subject behaviour. In terms of response times in an ambiguity identification task, the complexity of subject behaviour will be a consequence of the fact that by virtue of being naive, a subject is not necessarily oriented towards, or exclusively attuned to, the grammatically relevant properties of sentences.

It will be shown that the grammatical framework exploited by MacKay and Bever (hereafter, referred to as M&B) is inadequate to their purposes in two respects. First, although the theory is suggestive of differentiation among the categories of ambiguity, there exists no basis to predict that differential response must be exhibited in response times (other than by attribution to undefined notions of difficulty). Secondly, even the formal validity of the distinction between M&B's 'surface' and 'underlying' categories of ambiguity is in doubt.

A close examination of the manner in which M&B interpret their results reveals a number of implicit and unsupported assumptions. Most prominent among these are the assumptions that speakers' behaviour is a deterministic function of grammatical properties, that all speakers' (within a language community) share the same program, and that as a consequence it may be taken for granted that no systematic differences will exist among speakers.

There are several levels at which the M&B experiment may be criticized. In the most general terms, these are the degree to which their assumptions are consistent with the formal principles of transformational theory, the degree to which their procedures and analytical techniques are adequate to the experimental problem posed, and ultimately whether the essentially static rather than dynamic view of human choice behaviour is acceptabl . The introduction of the Choice Reaction Time paradigm is intended to provide an interpretive framework that is essentially dynamic. In these terms, the experimental problem may be re-expressed as involving two questions; first, whether the patterns of response correlate with levels of analysis or with types of information (i.e. is there a significant difference between surface and underlying ambiguities), and secondly, whether a single representative pattern of response could characterize the behaviour of all subjects.

An Overview of the Issues

The Notion 'Difficulty of Processing'

The M&B experiment consists simply of the observation of response latencies, given presentation of an ambiguous sentence taken from one of a number of formal categories of ambiguity. There is, however, no necessary relationship between the formal categories and potential for temporal discrimination among the categories. For interpretation, the MacKay-Bever experiment requires an ancillary theory of processing to relate the stimulus variable to the dependent variable. An adequate processing theory must be more complicated than a simple identification of levels of processing, since the formal categorization of the ambiguities is not based on mutually exclusive properties of the sentences. If temporal discrimination were revealed between lexical and structural ambiguities, such that lexical ambiguities were characteristically discovered first, this could not be explained by claiming that lexical information is processed first. If there exist distinct levels of processing, then no sentence can be comprehended until processing has been completed at each level.

The explanation of the relationship between experimentally observable temporal discrimination and categories of ambiguity has typically been couched in terms of the difficulty of processing that the presence of an ambiguity causes. The nature of this notion of difficulty is ambiguous, however, if one takes into account the M&B research and subsequent research based on it (Bever, Garrett, and Hurtig, 1973; Bever, Garrett, and Hurtig, 1976). Two interpretations of difficulty appear to be

1) difficulty due to processing mode,

2) difficulty due to stimulus properties.

The first interpretation seems to be associated with the idea of 'computational options', and in its strongest form could be said to assume that all associations between sentences and their meanings require equal time to process, variations in time result from the sheer number of possible associations that have to be examined before the meanings of an ambiguous sentence can be identified. The second interpretation deals with putative differences in the kind of perceptual 'rules' necessary to relate a sentence to its meanings as a function of the category of ambiguity to which it belongs. This interpretation assumes that the properties of some types of ambiguity inherently result in greater difficulty than for other types.

There would seem to be a confusion about the object of study. In one case, the effect seems to be attributed to speaker behaviour and in the other case, to the properties of the stimulus. Similarly, by neglecting to examine critically the manner in which ambiguities are categorized, an overly simplified impression of the relationship with speaker behaviour is given.

Grammatical Analysis and Message Processing

The character of formal analysis involves the identification of those properties that are necessary to draw distinctions between categories of sentences. Thus, with an ambiguous sentence a categorical distinction may be drawn between the grammatical structures found in the various interpretations. This could be said to be equivalent to a specification of the kind of information required to resolve the ambiguity. The sentence 'The boys ran out of boxes' may be resolved by noting the structural alternatives Verb-Preposition and Verb-Particle. This, however, overlooks the fact that the existence of the Verb-Particle description depends upon a fortuitous idiomatic interpretation of this specific lexical pair.

In an experimental situation, the very fact that subjects are naive may introduce a response orientation which conflicts with the formal basis on which the stimuli are distinguished. As noted above, semantically unique effects may be confounded with conventional and apparently regular grammatical structures. Such an effect might reflect a conflict between the differing terms in which grammarians and speakers evaluate sentences. For grammarians' purposes, the important features that sentences have are those that are maximally general in their relation to other sentences; one might characterize the speaker's point of view as having a closer relation to information theory, in which the most meaningful features of a sentence would be those that are least expected. This distinction could pose serious problems for an experiment dealing with ambiguity, and thus, having a unique involvement of meaning. The strategy alternatives could result in a confused response or in responses that might appear to be caused by something other than their true cause, the fact that a stimulus set must in any case be largely a grammatical and a semantically mixed bag will aid interpretation.

Meaning and 'Normal' Processing

It appeared in M&B to be an assumption necessarily associated with the notion of processing difficulty, that the general form of processing was invariant. Since the goal of processing is to extract meaningful information from sentences, one may ask whether or not our knowledge of meaning is compatible with the idea of an invariant form of processing. In very general terms, three types of theory of meaning can be identified (Harman, 1968),

a) Meaning as connected with evidence and inference,

b) Meaning as associated with ideas and messages,

c) Meaning as it involves social behaviour, speech acts, and communicative context.

Type (a), particularly when given a rigourous basis in logic or truth evaluation, is ill-equipped even to admit to ambiguity, as it appears to lead to paradox (Parsons, 1973). Type (b) is the level at which grammarians most commonly deal and is typically concerned with the issue of synonymy. The treatment of ambiguity at levels (b) and (c) seems to conflict; at least to the extent that sentences labelled ambiguous by the criteria of (b) may be disambiguated by context, while undecidability of meaning or nonuniqueness of meaning at level (c) stems from more than just conventional ambiguities (as in Quine's (1960, 1964) notion of 'vagueness').

In the absence of any comprehensive theory of meaning, it would seem fair to identify (a), (b), and (c) as independent aspects of the complex phenomenon of meaning. Likewise, there does not seem to be any principled basis on which to object to the idea that a speaker normally evaluates meaning with respect to all three levels. Mehler and Carey (1968) reported that the truth conditions of a sentence with respect to a picture display (a conjunction of levels (a) and (c)) have a systematic effect on response time, Carey, Mehler, and Bever (1968) report that there is an interaction effect between the truth conditions of the alternative interpretations of ambiguous sentences and response time (Carey, Mehler, and Bever, 1968). Similarly, evidence has been reported of positive correlation between the ability to correctly disambiguate sentences and a nonlinguistic dimension of field-dependent/independent (analytic/nonanalytic) cognitive style, associated with personality differences, (Lefever and Ehri, 1976).

Two points seem to emerge; first that the notion of the 'normalcy' of processing must be a relative notion, subject to context and goal of processing (which level of meaning is critical to the evaluation of the sentence's information); second, that to isolate the effects of grammatical differences, influences due to other levels of meaning must be controlled.

At this stage, the conclusions that may be drawn are limited to certain caveats concerning the expected source of influence that might be regarded as a determinant of speaker (or subject) behaviour and thus, concerning the limitations on the scope of inference entailed by the experimental observation of latencies associated with meaning-related decisions. The potential for confounding of influences associated with the levels at which meaning can be evaluated seems to be inherent even to the extent of admitting a component of subjective judgement in attributing ambiguity to any given sentence. The assumption that 'all things being equal' one kind of ambiguity might be expected to be more difficult than another requires a non-relative concept of normal processing, since otherwise the equivalence of circumstances would be meaningless. The question of processing is an empirical issue. The alternative to a strictly deterministic mode of processing, in which the categories of ambiguity correlate with observable patterns of response, is a flexible mode of processing which allows for a variety of strategy options. In the latter case, either the formal distinctions among ambiguities exercise an indirect control over the range of options (rather than direct influence over the pattern of response) or they are irrelevant to the determination of response.

Thus, the most fundamental criticism that might be levelled against the MacKay-Bever experiment is that by taking the apparent rigour of formal linguistic theory too seriously, they have been led to introduce the unjustified assumption that the stimulus items represent homogeneous classes, as indicated by behavioural response to them. Even less justified, they assume that the subjects' responses are uniquely regimented in some manner, corresponding to the principles of formal linguistic theory. In analytic terms, their experimental goal is to test the hypothesis that a difference in 'difficulty' exists between categories against the hypothesis that no difference exists; however, if processing is relative, then so is difficulty and a third hypothesis of conditional difficulty potentially exists. As a consequence, the rejection of the null hypothesis does not lead to acceptance of a unique alternative hypothesis.

The problems outlined above are general to any research involving behavioural response to the meaning of sentences. The specific issues relating to ambiguity must be considered in the light of a detailed examination of the way ambiguity is handled in formal grammars.

Transformational Grammars and the Classification of Ambiguity

Justification of Transformational Grammar

Transformational grammar (TG) has appeared to equip linguists to be able to characterize ambiguity as a coherent phenomenon, and ambiguity has likewise been used to motivate fundamentally the form of the grammatical theory. At least as presented in Chomsky (1965), TG has as one of its basic tenets the assumption that semantics is interpretive: meaning is independent of the representation to the same extent that a message is independent of the form of a code. There is assumed to be an interpretive function mapping meanings onto their representations, but the interpretive function must lie at least partially outside the province of a language specific grammar, containing as it must contextual, experiential and extralinguistic predicates.

In a manner bearing a close resemblance to the bi-uniqueness condition of structuralist phonemics, TG treats the interpretive function as requiring a direct correspondence between meanings and their representations and thus assumes that one of the adequacy criteria of a grammar should be to allow this correspondence to be as direct as possible.

Given the existence of classes of sentences that are judged to be mutual paraphrases, the relation between meanings and their representations would appear to be one-to-many. Given ambiguity the relation between meanings and their representations must also appear to be many-to-one. Thus, the first approximation to a meaning-representation mapping function would have to be many-to-many, despite the recognition of regular structural conditions that might be used to specify whether, in a given instance, a one-tc-many or a many-to-one mapping is required. From this rough characterization, the step from a taxonomic grammar to one involving representation of the systematic relations among sentences by an abstract and autonomous canonical structure appears to be motivated by the existence of paraphrase classes and ambiguities. Deep structure allows an intermediate stage onto which heterogeneous representations may be mapped and which may be mapped one-to-one onto meanings.

Independently of the relation between representations and their meanings, Chomsky (1965, 1966) has supported the notion of an autonomous deep structure by arguing that only at such a level could the logical relations among sentence elements be directly represented (i.e., relations of the type: 'subject of' and 'object of'). It was on this basis that M&B (1967) proposed a new category of ambiguity. Their claim was that two distinct types of structural ambiguity could be identified, one that could be resolved by assigning the appropriate surface structure, and another requiring specification of the deep structure.

The underlying structural level of sentences

> represents the essential "logical" relations between words and phrases ... Ambiguities at the underlying structure level involve neither a change in meaning of individual words, as in lexical ambiguity, nor a change in the apparent grouping of words, as in surface structure ambiguities, but only a change in the logical relations between words. (MacKay and Bever, 1967, p. 193, original emphasis)

This new category marks a significant departure of TG theory from the structuralist taxonomy of lexical, structural and phonological ambiguity. If the notion of logical relations is interpreted strictly, it may represent a different level of meaning from that commonly exploited in grammatical analysis. In the latter sense, a confounding may have been introduced, since while by definition each structural ambiguity will be associated with more than one deep structure, it is not clear that they will not also be associated with more than one distinct surface structure. Prideaux (1972) has argued convincingly that in the major class of examples identified as underlying ambiguities by M&B there would be distinct surface structure labelled bracketings. Thus, the formal validity of the category is in jeopardy. Garcia (1976), in a critique of M&B's research, has further noted that the inclusion of this new category results in non-unique classification of ambiguities (i.e., a structural ambiguity may meet the criteria of both surface and underlying ambiguity) and that the notion of an underlying category is misleading, in that it does not exploit the full range of properties represented by deep structure ("Why not ... relabel 'deep structure ambiguities' as 'ambiguities involving subject/object-verb relations'?" (Garcia, 1976, p. 202)).

Given the character of formal analysis, the issue concerning the underlying category of ambiguity primarily concerns the degree of generality involved. Any formal category may be represented in terms of less general subsets. However, the empirical finding that subjects respond differentially to the surface and underlying categories cannot be adequately answered in formal terms. Suffice it to say, at this point, that if the experimental findings are the prime justification for distinguishing between the categories and are also one of the only sources of evidence for a behavioural impact of the underlying level; then the argument is circular and it has not been established that the underlying level is the source of the experimental effect observed.

Classification and Judgement of Ambiguity

Apart from the question of the nature of the categories of ambiguity associated with a grammatical theory, there is a question about the appropriateness of regarding all instances of one type of ambiguity as being homogeneous. This problem is essentially that of the type-token relationship and was first raised with respect to ambiguity by Ziff (1965, 1967). Ziff's position is illustrated by the comparison of the following sentences,

(1) I saw the shooting of the hunters.

(2) I saw the shooting of the elephants.

It is Ziff's claim that in absolute terms (1) can be regarded as being ambiguous, but that (2) cannot be, given the discrepancy in potential for real-world reference between the two. He does not dispute that both share some potential for ambiguity, but rejects the assumption that ambiguity arises uniformly from the tokens. His claim, as paraphrased by Patel (1974), is that:

> ... grammars can only specify the types of morphological and syntactic structures whose tokens can be ambiguous in certain contexts. The crucial variable which determines whether a token of a given type of grammatical structure can be assigned more than one interpretation is its semantic state of affairs. (p. 32)

Gleason (1965) raises a similar pragmatic objection to premature identification of ambiguity in tokens. He draws on the idea of semantic 'redundancy' among the lexical items of a sentence. Thus, if 'chair' may refer to furniture or a professorial position, and 'table' may refer to furniture or data; then, in the following sentences,

(1) He gave the university a table.

(2) He gave the university a chair.

(3) He gave the university a table and a chair.

(1) and (2) could be termed ambiguous, but (3) could not.

In these terms, an issue is raised about whether ambiguity, per se, is a property of tokens or a property of classes of sentences. TG by its nature must treat ambiguity as a class phenomenon -- it is incapable of doing otherwise.

It is also clear that from a behavioural (judgemental) point of view this approach may not be appropriate. As a class, ambiguities identified by TG or other grammatical criteria, can be analyzed and classified in terms of their similarity, without the need to take account of their differences. This cannot be taken to guarantee a homogeneous behavioural response. Even the notion of sharing potential for ambiguity may require specification of a notion of 'grammatical similarity' in order that it be taken seriously. To argue that ambiguity is not primarily a token effect, it is necessary (a) that the presence or absence of ambiguity in a token of a particular type with potential for ambiguity must be associated with interaction between other general grammatical effects (e.g., in Ziff's example, ambiguity does not arise if the final NP is nonhuman), (b) that if the conditions relevant to two sentences of the same type are similar, then the alternative interpretations will also be similar, and (c) that the number of different kinds of interaction that might play a role in ambiguity will not be unlimited. The problems may be illustrated by examining the sentences below, each of which bears a superficial resemblance to Ziff's examples.

(1) I saw the dying of the day.

(2) I heard the answering of the question.

(3) I saw The Taming of the Shrew.

(4) I saw the painting of the Madonna.

It can be seen that the criteria listed above may be violated, these sentences share the same syntactic form as ziff's examples, but differ in type and number of interpretations but if the notion of sharing potential for ambiguity is to be meaningful, then these sentences must share the same potential for the same ambiguity.

Problems for the Interpretation of Experimental Results

The following general problems can be identified in the M&B experiment.

(1) Given the lack of explicitness about the precise relation between the stimulus properties and the dependent measure, the level at which stimulus properties might be influential in the process under observation must be regarded as an undecided empirical issue.

(2) The question of the potential for, and influences on, uniformity of subjects' responses is an empirical issue.

(3) The classification of ambiguity, both in clarifying the type-token relationship and distinguishing between varieties of structural ambiguity, is an issue that has not been resolved in the linguistic literature.

It seems clear that to expect to arrive at definitive conclusions about the behavioural impact of stimulus properties, the character of sentence processing, or the appropriateness of drawing certain distinctions between types of language material (with the implications that this might have for the interpretation of grammatical theory) is likely to be over optimistic, if following from the results of a single experiment.

The MacKay-Bever Experiment

Purposes and Procedure

The stated purpose of MacKay and Bever's study was,

> ... to study some of the factors influencing ease of perception of the interpretations of various kinds of ambiguous sentences ... (p. 193)

and the study was motivated by:

> ... the hope that by determining how subjects go about discovering the two meanings of a sentence, some insight could be gained as to how subjects go about discovering the one meaning of non-ambiguous sentences. (p. 193)

The latter motivation was given some elaboration in terms of a grammatical theory that might distinguish between lexical, surface, and underlying ambiguities. The distinctions were taken to imply ordered differences in the complexity of the categories, and this in turn was taken to suggest an ordered sequence of processing operations that a subject must carry out in order to resolve the ambiguity. Thus the second aim was restated:

> So the second purpose of this paper was to determine whether the *logical* order of processing of levels, postulated by linguists, corresponds to the *actual* or psychological order in which these levels are processed, on the assumption that ambiguities at those levels which are processed first should be discovered first, and ambiguities processed later, should be discovered later. (pp. 193-4)

M&B's experiment was organized in the following manner. Subjects were presented with 80 ambiguous sentences typed on index cards, and the experimenter instructed the subjects to respond 'yes' as soon as they had identified the alternative interpretations of each sentence. The index cards were handed to the subject one at a time, face down. As soon as the subject had turned the card over, the experimenter started a stopwatch which was stopped when the subject responded. The subject was required to describe the ambiguity immediately, and the time required for this was also measured. The two measures were termed 'Perception Time (PT)' and 'Verbalization Time (VT)', respectively. The index cards were presented to each subject in a different order.

Of the 80 sentences presented, 45 could be classified according to the syntactic taxonomy and were retained in the study. The remaining 35 sentences either contained other types of ambiguity or were discarded owing to extreme difficulty. Subjects were allowed a maximum of 90 seconds to respond; if they did not respond within that period, their PT was recorded as 90 secs. and the one interpretation that they could identify was noted. Of the total 1600 responses, only 15 of the interpretations were judged to be "ungrammatical, extremely idiosyncratic, or based on a misreading of the sentence" (p. 194). On the assumption that some of the interpretations were inherently more obscure or improbable than others, a measure of sentence bias was constructed by having the subjects report which interpretation they had recognized first. The percent frequency of the interpretation that was most commonly reported first was calculated from the resulting frequencies for the alternative interpretations.

Results

Results reported indicated significant differences between median PT's for each pair of categories of ambiguity, using a sign test. The rank of the categories in order of increasing median PT was lexical, surface, and underlying. Bias was found to interact with ambiguity type. In independent comparisons PT was found to be long if bias was extreme (either, high or low) in the lexical and surface categories. While for the underlying category, PT appeared to be short given initial recognition of the less likely meaning, but long given the initial recognition of the more likely meaning. The graph used to summarize this finding could be interpreted as indicating that for the underlying category, PT was a linear increasing function of bias; while for the lexical and surface categories, PT was a nonlinear (quadratic) function of bias.

The position of ambiguity in the sentence, the number of words in the sentence, the number of syllables in the sentence, and VT were all found to be unrelated to the ambiguity type temporal discrimination effect. Trends indicating generally decreasing PT as a function of trials were found in the data for each ambiguity type, suggesting a practice or learning effect. Some evidence drawn from responses to stimuli that included multiple ambiguities indicated increases in PT as the ambiguity occurred later in the sentence. The interpretation given was that lexical and surface material was processed left to right, but that a different manner of processing was indicated for underlying structures. The grammatical category of ambiguous elements was found to be related to PT, which was taken to be an effect of the relative salience of nouns and verbs as opposed to adjectives and adverbs. Surface structure complexity, as measured by the node to terminal node ratios of the surface structure trees of the stimuli, was found to be related to PT.

The results were interpreted as being generally consistent with the idea that sentence processing followed a pattern similar to the distinctions drawn by grammatical theory. An important deviation was, however, noted in the order in which the response to ambiguity type were ranked.

> Although the order of processing of levels postulated by linguists is the _logical_ order (underlying, derived, and then lexical levels), it is not necessarily the _psychological_ order in which people _actually_ process and understand sentences ... These results could be interpreted to suggest that the lexical level is processed first since ambiguity at this level was discovered first, and that the underlying structure level is processed last since ambiguities at the underlying level were most difficult to uncover. (p. 198)

Relative difficulty for interpretation was associated with the assumption that interpretation of ambiguous elements would be suspended until the remaining context of the sentence could provide a bias for one alternative interpretation; it was assumed that this process would be especially fast for lexical ambiguities, since they would be biased by semantic rather than grammatical relations.

A Critique of the MacKay-Bever Experiment

Analysis and Design

The most salient weakness of this study was the inadequacy of the statistical techniques used in the analysis. Given a dependent variable that is notoriously skewed, the authors opted for exclusively nonparametric techniques rather than the use of a normalizing transformation that would have allowed the use of parametric analysis. As a consequence, their analysis lacked power and more significantly, they were unable to test for crucial interaction effects.

The test used to determine the significant differences between median PT measure for ambiguity categories was the two-tailed sign test. The use of this technique requires that comparisons be made on twc variables at a time; thus, a test of joint independence on all three categories was impossible. Furthermore, the test assumes the signs of the differences between observations to be binomially distributed; this can only be the case if no subject by treatment interaction is present, and using this technique such an interaction cannot be tested (Siegel, 1956; Bradley, 1968; Sokal and Rolf, 1969).

The design used by M&B is essentially a repeated measures analysis of variance design. Given the assumption that no subject by treatment interaction exists, the validity of this design is contingent upon certain assumptions about the nature of responses to the ambiguity categories. These assumptions are related to the assumptions of compound symmetry and homogeneity of the variance-covariance matrices associated with the repeated measures ANOVA design (Winer, 1971). These assumptions, interpreted in terms of M&B's experiment, correspond to the requirements that (a) the between subject variance will be constant for all categories, because the same factors are involved in explaining subject differences in responses to each type of ambiguity; and that (b) subject strategies are regimented so that treatment differences will have constant parametric values, independent of subject differences. These assumptions are necessary for valid application of the sign test; but their compatability with the data is in question, given what M&B regard as evidence of differences in the character of processing between ambiguity types. Similarly, a subject by ambiguity type interaction, exactly the effect that the use of nonparametric techniques prevented the testing of, is precisely the effect whose significance is crucial to establishing the validity of a unique 'psychological order of processing', regardless of what that order might be.

The assumption of a unique psychological order is, of course, open to question in principle. It requires that decisions concerning the presence of ambiguity be independent for each putative level of processing. However, in sentences like 'They played with her present', it can readily be seen that the recognition of the full ambiguity is contingent upon the recognition of both a lexical and a syntactic component of the ambiguity. It is an empirical question whether an optimal strategy exists for uncovering all types of ambiguity, or whether a strategy which might optimize the discovery time of only one type of ambiguity might not have the effect of interfering with the discovery of other types.

M&B's treatment of the significance of their bias measure is similarly flawed. They report relationships between ambiguity category and PT, between bias and PT, and between bias and ambiguity category. Thus, the correlation between bias and PT can not be interpreted, since it confounds the relationships known to exist between those two variables and ambiguity category. Interpretation would require the evaluation of the partial correlation between bias and PT with the effect of ambiguity category held constant.

The inclusion of bias is a positive feature of this study, since it carries the implication that detection of ambiguity is more than just passive recognition. Moreover, the significance attributed to bias in the proposed processing model (in which interpretation of ambiguous elements is suspended until bias can be evaluated), in principle, seems to allow separation of type and token effects. The temptation to interpret the statistical relationship between bias and PT as revealing token effects, while the relationship between ambiguity category and PT revealing type effects must be avoided. This interpretation implies that the data are consistent with the assumptions of Analysis of Covariance, in which it would be assumed that bias would vary independently of ambiguity category. However, a differential bias by type effect is reported, which suggests that M&B's bias measures confound type and token effects with the consequence that an evaluation of the significance of these observations would be impossible without the use of Multiple Regression.

All criticisms of the treatment of 'subsidiary results', dealing with order of presentation, sentence length, syllable counts, parts of speech, etc., as independent influences on PT would follow the lines of the criticism of the treatment of bias. The joint evaluation of the influences of all of these factors necessitates the use of Multiple Regression, whatever the intercorrelations, and in its absence the reported results are rendered almost meaningless. As a case in point, Garcia (1976) noted that underlying ambiguities almost exclusively involved noun-verb relationships, and M&B report a part of speech effect which appeared to arise from the differential effect of nouns and verbs as opposed to adjectives and adverbs. It would be important to determine whether or not the part of speech effect could explain the category effect or vice versa, or whether the effects were after all independent.

In summary, M&B's analysis cannot lead to decisions concerning factors influencing the variability of PT. In each case, there seems to be a strong likelihood that multivariate analysis would have been more appropriate than the nonparametric techniques used. These considerations are apart from the questions of probable skewness of the dependent variable distribution, or heterogeneity of variance among subjects or stimuli. In any case, M&B's analysis is either invalid or has not been shown to be valid with respect to fundamental statistical and theoretical considerations.

Methodology

Similarly, serious methodological deficiencies exist, particularly as regards the task definition and stimulus preparation. The requirement that subjects identify ambiguities from among a stimulus set that is known to be composed entirely of ambiguous sentences will not necessarily require them to process these sentences in anything like a conventional manner. One need only posit a relationship between some lexical ambiguities and the metaphoric or idiomatic use of language to recognize that the task might sometimes simply involve scanning stimuli for superficial properties of phrases without any comprehensive analysis of the sentence as a whole. Nor would the requirement that subjects 'describe the ambiguity' control for this; since the subjects were only required to give an explanation of those aspects of the sentence that were relevant to the ambiguity. Clearly, nonambiguous sentences ought to have been included as a control.

Likewise, the grammatical construction of the stimuli was found to contain a bias. An examination of the stimulus materials reveals that roughly 21 percent of the lexically ambiguous sentences contain two clauses, while the corresponding frequencies for the surface and underlying categories are 13 and 60 percent, respectively. Thus, the experiment was likely to have been biased toward finding the longest response times for the underlying ambiguities, for reasons that are wholely independent of the putative nature of that category.

Interpretation

Finally, two aspects of the interpretation of the experiment are questionable. It is not clear that the notions of 'sentence perception', 'meaning perception', or 'perceptual rules' have any meaning, since perception can be regarded as direct response to physical sensation, and meaning, rules, and sentences are not physical entities. The use of these terms can only be regarded as an indication of the extent to which the authors regard a subject's behaviour to be regimented.

Most prominent among the assumptions that the authors brought with them into the study is the assumption of uniquely ordered levels of processing. The tentativeness of the assumption of uniqueness has previously been discussed; however, the assumption can also be criticized for its simplicity. When it is interpreted in association with difficulty, one is faced with the implication that sentences differ in difficulty along a single dimension. Thus, for instance, the combinations of obscure or common lexical material and simple or convoluted sentence structure would produce four classes of sentence each with a unique rank on the dimension of difficulty. It is not clear that such a generalized notion of difficulty is meaningful, though it is likely that the interpretation of temporal discrimination among ambiguities in terms of difficulty might be analogous to the claim that driving a car from A to B is technically more difficult than flying, because it takes more time. The question is empirical, and once again raises the question of whether or not the response space should be multidimensional.

The Character of the Dependent Variable

Thus far, discussion has dealt with technical and theoretical aspects of the experiment per se, not with broader questions of the use and interpretation of latency as a dependent measure or with the implications of more general issues from experimental psychology.

Essentially, M&B's procedure is a test of what must be assumed to be a deeply engrained discriminative ability of English speakers. But, while the existence of a general ability might be taken for granted, the nature of the discriminative process cannot be. As Kling and Riggs (1971) noted:

> It should be apparent that discrimination or differential responsivity, is the obverse of generalization. When an organism responds in the same way, or similarly, to two or more stimuli it may be said that generalization exists among the stimuli -- for that organism executing that response. Discrimination involves the absence or dimunition of generalization ...
>
> Because an organism's discrimination is always assessed through the measurement of some particular response property, discrimination performance can be at the same time both excellent and poor, even with respect to the same set of stimuli. Thus, a subject might choose an incorrect stimulus as often as a correct stimulus, but take longer to choose the incorrect one. 'Percent correct' would show no discrimination, but 'latency' would. Thus, discrimination capacity of the subject must always be defined in terms of the response measures that are used. (pp. 747-8).

Two crucial points, bearing on the M&B experiment, are made in the characterization above. First, one is led to ask what the relationship might be among stimuli for which no temporal discrimination is found. If responses to lexical ambiguities are homogeneous, the implied generalization among lexical ambiguities cannot be construed as a relation among properties of this kind of ambiguity. Rather the generalization must imply a common strategy for identifying this form of ambiguity, since in terms of semantic or other features, lexical ambiguities are an inherently heterogeneous class. Thus, the most direct inference leading from the finding of temporal discrimination between lexical and other forms of ambiguity would be that the processing mode or strategy differs.

Secondly, since the relation between simple stimulus properties could not be the basis for stimulus generalization and differing processing modes are implied, then differences in the accuracy of processing between modes is likely. Thus, if discrimination is found in latencies, then it would likely be found in errors of identification as well. However, M&B's experiment is reported as being essentially error free. This being the case, there is no discrimination in accuracy of response and this finding conflicts with the idea that ambiguities are being ranked on a dimension of 'difficulty'.

If the assumption of a series of distinct levels of processing is in the least tenable, then it would follow that if processing time were cut short an increased likelihood of error in identification would result. This was the basic assumption underlying an experiment reported in Prideaux and Baker (1976), in which subjects were presented with both ambiguous and nonambiguous sentences on a computer controlled CRT screen. Stimuli were displayed for a fixed period of seven secs., and subjects were instructed to decide whether or not each stimulus was ambiguous during that time. There were 32 sentences in the stimulus set, composed of eight nonambiguous and eight each of lexical, surface, and underlying ambiguities adapted from M&B's

stimulus list. The sentences had been balanced for length, and each had two clauses. Within the set of ambiguous sentences, half had the ambiguity in the first clause, half in the second clause. An ANOVA on the errors in identification revealed significance only for the ambiguity category main effect. A subsequent series of t-tests for correlated means revealed significant differences between the lexical and surface, and the lexical and underlying categories. There was no significant difference between the surface and underlying categories.

The obvious conclusion is that discrimination can take place on the basis of error in identification, and that the pattern of three way discrimination among the categories does not carry over from M&B's latency study. Moreover, comparison of M&B's error free study with the error rate of this rapid response experiment is suggestive of a speed/accuracy trade-off in the response process.

The Relation between Stimulus Categorization and Latency

The Choice Reaction Time Paradigm

The major theoretical problem encountered thus far has been the interpretation of the notion of difficulty in processing. Attempts to predict temporal discrimination in terms of the nature of stimulus processing have led to ambiguous conclusions. An alternative would be to attribute a functional relationship between latency and the strength of association between the stimulus and the appropriate response. One could in this case posit an inverse relation between associative strength and difficulty. In cases where the response task is well-established, so that little learning would be anticipated as one would expect with language behaviour, interpretation of processes as inferred from latency data falls within the framework of the CRT paradigm.

Smith (1968) cites a variety of studies in which the stimuli were composed of complex perceptual features, and in which verbal responses were chosen on the basis of a conceptual association between those features and the response set. In such tasks, it was frequently found that response time varied directly with the size of the stimulus set, consequently with the number of distinctions between stimuli.

Examples of the kinds of tasks involved in this research include colour naming, association between colours, and naming letters, faces, animals, geometric symbols, etc. (Catell, 1947; Woodworth and Schlossberg, 1954; Morin, Konick, Troxell, and McPherson, 1965). These studies appear to indicate that the effect of the number of disparate elements in the stimulus set diminishes as the associational strength between the specific stimuli and the appropriate responses increases. Broadbent and Gregory (1962) investigated sorting-time for a set of cards with familiar place names printed on them; the task required either sorting on the basis of typeface characteristics or associations between the place names. No significant difference was found between these two types of task, except when the required association between places conflicted with the subjects' expectations. They concluded that "when a set of stimuli are grouped together by common usage the response time is ... independent of the response/stimulus ratio" (p. 1315).

In an example of a numeral-naming experiment, average response time for naming a subset of the numerals chosen as stimuli decreased when the appropriate response to numerals that were not in the subset was 'no' rather than the numeral's name (Forrin & Morin, 1966), this was interpreted in terms of "generalized response competition". It is consistent with a finding for the a priori probability of the stimulus (Dillon, 1966), and the finding, in a motor response task, that "a measurable portion of the reaction time interval is consumed by processes associated with the inhibition of competing incorrect alternative response" (Kornblum, 1965, p. 55).

Thus, the psychological literature, in general identifies three main components of the reaction time measurement, each of which could be expected to contribute to the difficulty of stimulus processing. The major contributor is the size of the stimulus set; this could be likened to a general indication of the difficulty of the task. Secondly, a high associative strength can facilitate the choice of the correct response to a given stimulus, or competition among response alternatives be detrimental to rapid response. In any event, the basic idea underlying the CRT model is that response time can be related directly to the information content of stimulus presentations, where information is measured in terms of the number of distinctions necessary to distinguish the entire stimulus set (Welsford, 1960; Hick, 1952).

CRT Processing Models

In order to account for the variety of empirical results obtained from CRT experiments, three basic forms of processing models have been specified. The major difference between the models is that of global vs piecewise processing of stimulus information. This distinction is expressed in terms of the distinction between Template Matching (TM) and Feature Testing (FT) models. A second overlapping distinction rests on the question of whether or not a decision is reached in a deterministic fashion, or alternatively whether a decision is made with only a certain probability of being correct. This allows the possibility of a Statistical Decision-making (SD) model, which may bear similarities to either the FT or TM models. Each model assumes that the process can be divided into three stages, though not necessarily altogether disjoint stages: Stimulus Preprocessing, associated with perceptual recognition, Stimulus Categorization, associated with the identification of the stimulus, and Response Selection, involving selection of the appropriate response contingent upon the indentification of the stimulus.

The TM models assume that the most important stage of the CRT process is that of stimulus preprocessing in which internal representations (replicas) of the stimulus are constructed. Having constructed the replica, the subject is able to compare this with templates stored in memory in order to identify the stimulus; thus, the TM model is basically a memory search. A number of different detailed predictions about response time may be made, dependent on whether the time required for a comparison is a constant or a random variable, whether comparisons are made in sequence or simultaneously, and whether the process is terminated when a match is found or when the total number of possible comparisons to be made is exhausted. In any case, a general prediction is that a direct relationship will exist between mean reaction time and the number of stimulus alternatives set forth in the task.

The FT models assume that identification of the stimulus may be achieved by adverting to certain specifiable properties which the stimulus may or may not have. Underlying this assumption is recognition that for a given task all of the information carried by a stimulus need not be relevant to categorization of the stimulus. Thus, in contrast to the global processing of the TM model, the FT models have more in common with pattern detection and do not rely as heavily on accessing memorial information. For the purposes of making predictions about reaction time, it must be assumed that the tests for relevant features are carried out in sequence. If the tests are considered to be a sequence of dichotomizing tests on 'n' relevant features, then the number of stimuli that could be distinguished on the basis of these features is n^2 and the number of tests required to identify one of the stimuli will be the log of the number of stimuli. Thus, assuming that the time required for a given test is a constant or a random variable with constant variance, reaction time for FT models will increase proportional to the log of the number of stimulus alternatives.

The general differences between the FT and TM models are that TM is in conception simpler and implies a more direct relation between the complexity of the stimulus and CRT than does FT. On the other hand, FT models are potentially more flexible and could be adapted to account for learning or practice related effects. It may be argued that in relation to identification of ambiguity, the only way in which dichotomizing tests could be taken seriously would be in terms of a transformational model, whereas a TM model could involve extra-linguistic factors.

The basic principle underlying the Statistical Decision-making model (SD) is that if one conceives of processing as repeated sampling of stimulus information, then the greater the number of samples taken the more accurate the decision is likely to be. However, since sampling requires time, an increase in accuracy will result in a decrease in speed. The terms in which SD models are expressed are generally related to sequential statistical procedures and the evaluation of the log likelihood ratio (Stone, 1960). It is necessary that the subject have clear alternatives of response, although the stimulus information that he samples may be fallible. It is assumed that the subject can form distinct hypotheses to be evaluated and can set criteria which will represent the limits necessary for acceptance of a hypothesis, or in other words, stimulus categorization or identification. The criteria are independent for each stimulus or response alternative, and can be biased differentially for each hypothesis (associative strength), and changed with experience (learning). Thus, the major feature that distinguishes this general form of model from deterministic counterparts, is the existence of a systematic relationship between error rate and reaction time.

Ambiguity and the CRT Paradigm

The implications that research into CRT have for ambiguity research are various, but foremost is the characterization of what reaction time measures. The CRT literature reveals that even in straightforward motor tasks, the nature of the response causes a measurable effect on reaction time. It is not necessary to look any farther to see that the interpretation of reaction time to ambiguities in studies involving a completely open-ended response can not be interpreted in terms of complexity of processes leading to comprehension. If a subject's task requires that he have the form of his explanation of the ambiguity in mind before responding, then reaction time measures a complex stimulus identification process and a complex response selection process. That is, both the processes of comprehension and of expression are being measured. If variations in the associative strength contribute to reaction time, then so would response competition, not simply between competing incorrect alternatives but also among the many ways of expressing the correct response. If the CRT literature is taken as seriously as it should be, then the only way in which any reliable indication of variations in reaction time associated with comprehension processes could be observed would be strictly to limit the response alternatives. This could be accomplished by requiring the subject to identify the kind of ambiguity that a given stimulus represents; or given a mixture of ambiguous and nonambiguous stimuli, the subject could be required to identify the presence of ambiguity. Thus, responses could be limited to either 'lexical', 'surface' or 'underlying' or to a simple 'yes' or 'no'.

The notion of difficulty associated with the CRT literature is not a notion of relative processing difficulty between populations of stimuli. The problem of temporal discrimination does not arise, since CRT theory seems to assume that for a given subject reaction times will largely reflect overall task difficulty. One possible way in which temporal discrimination could be accounted for without introducing an ad hoc "fudge" factor would be to allow for more than one form of stimulus information processing. Thus, for example, lexical ambiguity might require an appeal to memory, while structural information might be more efficiently handled by a feature testing model. In such a case, it would be difficult to formulate a model which would be compatible with temporal discrimination between surface and underlying ambiguity, since the distinction is so closely tied to the grammatical framework.

As a final point, it should be noted that in CRT experiments the overall task difficulty is tied to the average information carried by the stimuli. Information in this sense is a relative concept, in which bits of information can be used to distinguish classes of stimuli from each other. In an open-ended ambiguity task, each ambiguity represents a unique class. One can not expect a subject to anticipate any of the stimuli, and thus, if one can not expect that the subject has assigned subjective probabilities to the occurrence of stimuli, then it is not possible to talk meaningfully about the average information or difficulty. From the information theoretic point of view, the amount of information required by the subject is the amount necessary to distinguish all possible ambiguous sentences from each other. In this sense it can be seen that a task which requires the identification of the unique meaning-representation associations of ambiguous sentences lays itself open to maximum inconsistency of response, and the maximum variability due to unique semantic and subjective influences. Furthermore, the notion of difficulty under examination has to be one in which categorical differences in difficulty are hypothetical. If the long history of CRT theory has any substance to it, it must imply that if subjects are not aware of categorical differences in the stimulus set then the experiment can not be said to be properly controlled.

Psycholinguistic Research on Ambiguity

Because of the unique position of transformational grammar with respect to ambiguity, most psycholinguistic research on ambiguity has been flavoured by its grammatical principles. A theoretical position which has been given much attention may be characterized as "a general theory of sentence perception ... which emphasizes the processes of clause segmentation, recoding clause by clause, and the use of direct mapping rules to assign internal relations to surface sequences" (Bever, Garrett, and Hurtig, 1973, p. 285). It is in this sense that the speaker/hearer is said to compute the meaning of a sentence (Garrett, 1970). The fact that certain ambiguities might require that such computation reach a certain stage before the ambiguity can be identified is the theoretical basis for accounting for temporal discrimination among types of ambiguities. There are two basic weaknesses to this characterization. First, if sentence comprehension processes proceed in stages, then all stages must be analyzed before the sentence can be understood. Thus, all sentences regardless of whether they are ambiguous or what type of ambiguity they might contain must be processed to the same level in order to be comprehended. Discrimination must now be explained in terms of the fact that the ambiguity was hypothetically located at one level rather than another; that is, it must be explained in terms of the differences in the nature of the level, the kind of processing involved, or the nature of the information processed at that level. Second, insofar as the characterization draws on grammatical entities (e.g., clauses, surface and underlying structures), then it must be consistent with the grammatical theory that is the source of these concepts. However, the grammatical principles are violated, unambiguously, in the stipulation that "within each perceptual unit [clause], the underlying structure relations are assigned by direct projection from the lexical sequences" (p. 272). It is precisely one of the motivations for an underlying structure level, that it can be used to represent systematic relations between interdependent clauses.

Many of the experiments dealing with differential response to ambiguity have used as dependent variable the time required to complete potentially ambiguous sentence fragments cr to produce a sentence following an ambiguous stimulus sentence (MacKay, 1966; Mistler-Lockman, 1972; Bever, Garrett , and Hurtig, 1973). These studies are particularly sensitive to the criticism that they primarily involve observation cf response selection and can not be given interpretation with regard to comprehension processes. It is noteworthy that no theoretical basis exists for predicting discrimination associated with processes of expression, unless the process is conceived of as making use of transformational grammatical principles operating in reverse.

However, Foss, Bever, and Silver (1968) devised a semantically oriented task which circumvented the need for the subject to respond with his own characterization of the content of the stimulus. Ambiguous and nonambiguous sentences were displayed to the subject along with a picture which might represent activities described by the sentence. For each nonambiguous sentence in the stimulus set, there was only one sentence-picture pairing that was appropriate semantically, for each ambiguous sentence there were two. The subjects were required to indicate whether or not randomly paired pictures and sentences were semantically related. The results yielded discrimination between ambiguous and nonambiguous, and between lexically and non-lexically ambiguous stimuli in terms of response times for correct responses only. It is of interest to note that while the experiment meets the task definition criteria for a CRT experiment, the pattern of discrimination reported in M&B (1967) is contradicted. The results are consistent with the only other study known to meet the CRT criteria, Prideaux and Baker (1976).

The Foss, Bever, and Silver results were interpreted as emphasizing the importance of bias: "Ss typically assign only one immediate interpretation to an ambiguous sentence. Only if that interpretation is found to be correct does S reinterpret the sentence" (p. 306). However it was found that bias was not accurately estimated by pretest evaluation of the relative saliency of the alternative meanings; a more accurate assessment was obtained by questioning the subject about his immediate expectations. This latter procedure corresponds to the procedure used by M&B, and casts doubt on the inherent, normative relationship between the stimuli and associated bias measures. If bias is not a function of the expected semantic state of affairs of a sentence, but is a function of semantics and the disposition of the subject; then it can not be used to explain subject behaviour.

In general the existence of temporal discrimination among ambiguity types, as a phenomenon, can not be dealt with by current characterizations of the nature of sentence processing. It may be accepted that response time discrimination exists, but no theory of sentence processing necessarily implies that this must be the case. In order to account for the phenomenon, it is not sufficient merely to establish that a difference among types exists. The theory must imply that for soundly based reasons, discrimination will arise in *response times*.

The general theoretical treatment of ambiguous sentence processing has not reached the stage of identifying the major sources of variation, let alone the structure of that variation. A number of pertinent issues have been raised, however. One issue is whether a subject analyzes both interpretations of an ambiguity and chooses between them, or analyzes one first (presumably the most heavily biased) and, on being forced to reconsider, produces the alternative interpretation. Garrett (1970) notes that only in the former case would the nature of processing ambiguous sentences differ generally from the processing of nonambiguous sentences. The issue has been characterized as distinguishing between an Exhaustive and a Unitary model of processing. The Unitary model is in part motivated by the so-called 'Garden Path' phenomenon, where (as in Lashley's (1951) example) conditions similar to Gleason's (1965) 'semantic redundancy' lead to standard expectations about the interpretation of ambiguous elements. It is not clear that similar effects must carry over in experiments in which the subject actively searches cut ambiguity, but the common assumption that processing is deterministic has led to the expectation that it would. The Exhaustive computation model requires an ancillary theory to account for the final decision to opt for one interpretation or another. One possibility, discussed by Bever, Garrett, and Hurtig (1973), is the Perceptual Closure model. It predicts than an interpretation will be assigned independently to the first clause analyzed and the interpretation of the subsequent clauses must be consistent with it. The usefulness of this model is minimal, in that it does not account for the manner in which the choice is made, except by vague reference to perceptual rules. An alternative ancillary theory involves incorporation of the Perceptual Suppression process, frequently encountered in the theory of visual perception of ambiguous figures. In this model, the importance of bias is made most evident. MacKay (1970) outlines the basic assumption as being that "in order to perceive one meaning of an ambiguous sentence, the other meaning must be suppressed and the time to suppress a meaning varies with the salience of that meaning in the context of a sentence" (p. 86). The measure of salience is associated with the bias of a sentence, which measure was noted before as potentially involving speaker judgements. This model seems to have similarities to the response competition effect reported in the CRT literature, and in that sense reveals a weakness in the modelling orientation. There seems to be a basic confusion between an interpretation and another sentence. If comprehension is demonstrated by uttering a paraphrase, this does not imply that comprehension is regarded as the creation of paraphrases. It may be that the analogy to visually ambiguous figures has deceptively obscured the process transition from sentence comprehension to response selection.

It is interesting to note that the Perceptual Closure model carried with it the assumption that processing of ambiguous and nonambiguous sentences involved the same mechanisms; while the Perceptual Suppression model introduced a parameter which would specifically complicate the analysis of ambiguous sentences (unless competition among alternative interpretations of greater or less semantic differentiation is posited for nonambiguous sentence processing). MacKay (1970) raises the question of interpretations being entertained with varying degrees of likelihood. This issue, termed the Tentative Decision Issue, in introducing aspects of signal detection theory would invert the usual assumptions. There is no necessary reason for the normal process to be the computation of the most likely meaning rather than the elimination of unlikely interpretations. In such a case, processing ambiguous sentences would be the standard, and it might be expected that a unique interpretation, in the case of nonambiguous sentences, may be hard to achieve.

In summary, it may be said that few experiments involving differential response to ambiguity have managed to control the open-ended response selection processes which almost inevitably contaminate research in this area. Furthermore, no theory of sentence processing has incorporated predicates that would predict discrimination among types of ambiguity in terms of response time. No theory has adequately dealt with the potential variation in accuracy of response, or the potential trade-off between speed and accuracy. No clear or unambiguous characterization of the nature of relative difficulty exists.

The current linguistically oriented theories have been so concerned with the minutiae of stimulus differences (token effects) as to be predisposed to assume an essentially passive, uniquely determined response on the part of subjects. The whole area of general task complexity, subjective bias, and the range of potential sources of heterogeneity of responses have been neglected. It would seem appropriate to revise the direction of research in this area to take into account the human subject, in a manner similar to the practices of CRT theory.

In the following chapters, two experiments dealing with ambiguity will be reported. These experiments will attempt to replicate M&B's results, but with the following provisions: (a) that the overt subject responses will be reduced to a simple yes/no identification of the presence of ambiguity, (b) that nonambiguous stimuli will be introduced as a control, and (c) that where ever necessary the appropriate multivariate analytical techniques will be used. Chapter Two will deal with the experimental design and the methods used. The direct results of statistical analysis will be reported in Chapter Three. The discussion, presented in Chapter Four, will attempt to reconcile the findings with

M&B's results, with other literature on the subject, and to draw directly on the implications of CRT theory. In the conclusion, Chapter Five, the implications of this study will be summarized and note will be taken of implications for further research.

CHAPTER TWO

METHOD

The attempt to replicate and extend the M&B experiment was undertaken in two parts. The logical distinction between type and token effects suggested that observations of ambiguity identification response times ought to be made independently of the estimation of token bias. As a consequence, the experimental methods will be outlined as if for two experiments, although ultimately bias will be treated as a covariate of response times, rather than being analyzed independently.

Experiment 1 dealt with subjects' response times to ambiguous sentences. Its basic aim was to attempt to estimate the effects on response time due to general semantic heterogeneity, subject differences, clause effects, and ambiguity type. Experiment 2 involved collecting judgements of the bias in the stimuli used in Experiment 1.

Experiment 1

Subjects

The subjects were 42 undergraduate students from introductory psychology classes. Their participation in the experiment was part of their class requirement, in lieu of a lab, and counted 5% towards their final grade. They were given a free choice from among a variety of experiments to participate in, and given the negligible contribution to their final grade were free to choose not to participate in any.

By the nature of the task, subjects had to be run individually. In order to organize this, booklets containing session times were made available to the classes and subjects signed up for the times most convenient to them. The sessions were run over a period of two weeks, with subjects being assigned to experimental conditions in the order in which they appeared. As might be expected, a certain amount of confusion took place with some subjects mistaking their appointment times and others not making an appearance. As a consequence, although the subject set was balanced across experimental conditions and between the sexes, there was an uneven distribution of the sexes across the experimental conditions.

Apparatus

Stimulus presentation was computer controlled, and for this purpose a program was written for the Linguistics Department's PDP-12 by Dr. Anton Rozsypal. The program incorporated the following features:

(a) Individual stimulus lists and assembled versions of the presentation program were stored on LINC tape.

(b) Each time the program managing any one of the stimulus lists was loaded, a randomizing routine was initiated, resulting in a unique stimulus order for each session.

(c) For testing purposes the stimuli were displayed on a remote cathode ray screen; external lines to the computer allowed the experiment to be run and controlled by the experimenter in a room away from the computer.

(d) The presentation of each stimulus involved the following sequence of events:

i. Initiation was indicated by a three second warning display of asterisks.

ii. The stimulus display followed immediately and the clock was simultaneously initiated.

iii. Timing ceased if either one minute elapsed or the response button was pressed.

iv. If the one minute time limit was exceeded, the stimulus sentence was removed from the screen. Otherwise, the sentence remained on the screen until the identification was registered as correct or incorrect and the next presentation initiated.

(e) Throughout each session, the computer kept a running account, printed on the teleprinter, of the trial numbers, the stimulus item associated with each trial, an indication of the correctness of response or whether the time limit had been exceeded, and a running total of the number of correct responses up to and including that trial.

Materials

The examples of ambiguity used in this experiment were drawn from M&B's list of stimulus materials, subject to an informal survey of the Linguistics Department members' impressions of the relative obscurity and acceptability of each example. The basic stimulus set from which presentation lists were drawn comprised 42 distinct sentences. Of those, 24 sentences were ambiguous and 18 were nonambiguous. Within each of the categories (Lexical, Surface, and Underlying) of ambiguity, there were eight instances representing the most uniformly acceptable examples from M&B's materials.

Each of the examples drawn from the M&B list were modified to meet the following criteria:

(a) each sentence was to have two distinct clauses;

(b) each sentence was to consist of 12 words.

The latter condition was relaxed when it was judged that strict adherence to the criterion would result in superfluous padding with adjectives, or in a stylistically awkward construction. As a consequence, two instances were left with 11 words and three with 13 words. In any event, since the purpose of maintaining a constant sentence length was to attempt to control for basic variations in reading time, it was not felt that a margin of one word would be unacceptable.

Initially, 24 nonambiguous sentences were constructed, so that for each ambiguous sentence there was a nonambiguous sentence with a corresponding surface syntactic pattern. On the basis of a subjective judgement of stylistic acceptability, six nonambiguous sentences were discarded.

Six presentation lists were then constructed from the basic stimulus set, in the following manner:

(a) each presentation list was to contain six instances of each of the categories (lexical, surface, underlying, and nonambiguous), totaling 24 stimuli per list.

(b) for each list containing a specific sample of stimulus sentences, there was to be a corresponding list which would be identical except that the relative positions of the clauses would be reversed.

(c) for the categories of ambiguity, maximum differentiation between the lists was to be achieved by ensuring that each pair of lists had only four items in common, but that the four would be unique to each pairing.

(d) each list containing a specific sample of ambiguous stimuli was to have a unique set of nonambiguous stimuli.

These conditions establish three fully crossed factors: List, representing each of the three pairs of semantically identical samples of stimuli; Position, representing the alternative orders of the clauses; and Type, representing the stimulus categories.

It was realized after the experiment was underway that due to an oversight condition (c) had not been fully satisfied. Within the surface category, the Lists were found to have three instances of ambiguity in common; one pair of Lists had five ambiguities in common, and consequently, one ambiguity appeared in a single List only. This flaw had potential for introducing an unintended bias in the differences between the Lists.

Procedure

The instructions given to the subjects (see Appendix A for the full text) explained that the stimuli were composed of both ambiguous and nonambiguous sentences, but that there were more ambiguous than nonambiguous. They were told to respond 'yes' or 'no', the moment that they had decided whether or not a given stimulus sentence was ambiguous. The experimenter then pressed the response button to record the time taken to reach the decision. If the response was 'yes', the subjects were required to write a sentence corresponding to each interpretation they saw, in a booklet provided. The sentence remained on the screen for reference during the paraphrasing, and was removed only when the experimenter pressed the buttons indicating a correct or incorrect response. At this stage, correctness of response represented correct categorization and not correct interpretation or paraphrase.

Design

The basic analysis of variance design was a five-way factorial (List, Position, Type, Subjects, Replicates) with List and Position fully crossed with Type, with repeated measures on each subject across the levels of Type, with subjects nested under List and Position, and with six observations within each cell. The data structure is illustrated in Figure 1 .

The basic experimental hypotheses under test are as follows:

i. Main effects:

List: Are response times to semantically differing sets heterogeneous? Care must be exercised in interpreting this factor, since the unbalanced distribution between sexes, and the nesting of subjects opens the way for confounding with subject effects.

Position: Are there systematic differences in response to semantically identical material differing in terms of clause order? This position effect will not directly represent a clause effect, since each level of Position contains items with the ambiguity in both the initial and final clauses. This was necessary in order to prevent learning or formation of clause related strategies which ignore the irrelevant clause.

Type: Are there significant differences in the speed of response, dependent on stimulus type?

Subjects (nested within List and Position): Are there significant differences between subjects presented with the same stimulus material?

ii. Interactions:

S x T: Are there significant differences in the pattern of subjects' responses to ambiguity types, that might represent differing strategies?

L x T: Since there is no semantic correspondence between levels of Type, differences in response associated with true differences in semantic material would be associated with this effect.

P x T: Since variation in clause position for nonambiguous stimuli ought to have no effect on the detection of ambiguity that could be attributed to clause by clause processing; a clause effect which might imply involvement of clause order in processing will be associated with the significance of this interaction.

iii. The interpretation of all other effects must be contingent upon the pattern detected by a posteriori tests.

A note of uncertainty concerns the appropriate structural model for the analysis of variance. There can be no question that List, Position, and Type are fixed effects. The question of whether subjects can be considered fixed or random, depends on whether or not their responses are considered to be highly conditioned by the experiment. If a variety of strategies is indicated, then the subjects might be said to have adapted their responses to the experimental situation. In that case, the population would be fully represented by the subjects and they could be treated as a fixed effect.

Experiment 2

Subjects

The subjects were a total of 105 undergraduates, taken from a number of Introductory Linguistics classes. The subjects were tested prior to having been introduced to formal syntax or any sentence level grammatical analysis, and thus, could still be considered naive. Of the students who participated in the task, the responses of 10 were discarded as being either non-native speakers of English or speakers of non-North American dialects of English. As the classes were overwhelmingly composed of females, it was decided not to distinguish between the responses of the sexes.

Materials

All of the ambiguous stimuli from the response time task were assembled. Only one of the alternative clause orders was selected, and booklets were prepared in which each ambiguous sentence was listed together with a pair of appropriate disambiguating sentces or phrases. Thus in the booklet, each ambiguous stimulus was represented in a form similar to the following:

1) The car kept stalling while they discussed
   their problems with the mechanic.
   A. They had problems with the car.
   B. They had problems with the mechanic.

The booklets contained the ambiguous stimuli in one order only, and all of the disambiguating phrases were identical for a given ambiguity in all booklets.

Procedure

The subjects were given the booklets during a class period, and were verbally instructed to read each ambiguous sentence carefully, then to read the disambiguating phrases, and to circle A or B depending on which interpretation occurred to them first, or was judged to be the more natural interpretation of the ambiguous sentence.

The data were scored by counting the number of times that A and B were circled, for each ambiguity. The total of the more frequent of the two was then divided by the total number of responses. This score, representing the proportion of responses associated with the preferred interpretation, was taken to be a measure of the bias of the ambiguity. The bias score would be near 1.0 in cases where one interpretation was highly implausible, and near 0.5 in cases where both interpretations were equally plausible.

CHAPTER THREE

RESULTS

Preliminaries

The results are divided into two general areas of interest; the first being the examination of response times in accordance with the experimental design, and the second being the joint evaluation of covariates of response time.

The analysis of response times took the form of individual analyses of variance of the responses under each stimulus category, and two multivariate analyses of variance, one including all stimulus categories as independent variables and another involving only categories of ambiguity. Two conclusions are derived; first, that there is no empirical evidence that subjects responded differentially to the surface and underlying categories of ambiguity, and second, that the major source of variance was the differences among subjects. No significance was detected due to variation in clause order or of semantic material. Subsequently, an attempt was made to determine the general properties of subject strategies. This involved a discriminant function analysis incorporating lexical ambiguity, structural ambiguity, and nonambiguous sentences as classification groups, while individual subjects were treated as independent sources of variation. It was found that no overall pattern of response could be attributed to the subjects as a whole, that the significant results obtained in the analysis of variance were associated with a distinct subset of subjects who responded differentially (termed 'discriminators'), and that these discriminators' strategies were best described in terms of the patterns of response of four distinct subgroups. Finally, it was noted that none of the subgroups of discriminators responded differentially to all three stimulus classifications; in each case the pattern of response was essentially dichotomous, although a different dichotomy for each subgroup.

The analysis of the covariates of response time was accomplished by means of a set of related multiple regressions. The goal of these regressions was to evaluate jointly the potential contribution of variables representing (a) learning or practice effects, (b) token effects (semantic bias), (c) subject bias (propensity to respond 'yes' or 'no'), (d) sex effects, and (e) errors in categorization or paraphrase, together with classification variables representing the results of the analysis of variance. The results obtained indicated significant independent contributions from three types of variable: (a) subject related variables, comprising sex, subject bias, and strategy group, (b) error related variables, including error in categorization, error in paraphrase, and an error in categorization by ambiguous/nonambiguous interaction, and (c) stimulus type effects, including a simple ambiguous/nonambiguous effect, and a practice effect on the lexical/structural ambiguity dichotomy. The overall proportion of variance explained by the regression was low (about 15%) suggesting either a very noisy underlying process, or failure to identify important sources of predictable variation. If the latter were the case, it would seem likely that given the exhaustive treatment of error and type related effects, significant aspects of inter-subjective difference remain unidentified.

Response Time Analysis

Prior to subjecting the data to analysis, two transformations were applied to them. In the first instance, the response time observations were found to have a highly skewed distribution and in order to both normalize the distribution and stabilize variance, the data were subjected to a log transformation. Subsequently, an analysis of variance revealed only subject differences as a significant effect. However, these results were not given credence, as Bartlett's test revealed highly significant heterogeneity of variance among subjects, chi-square (41) = 179.8, $p < 0.0001$. On the assumption that basic regularities in response could be masked by simple subject differences in mean and standard deviation, the data for each individual subject were standardized. The variance-covariance matrices associated with the stimulus categories were examined, in light of this transformation, in conforming to the assumptions of repeated measures analysis of variance. The conditions of compound symmetry and homogeneity which would justify the pooling of information required by repeated measures analysis (Winer, 1971) were found to be violated. Following Winer's suggestion that in such situations "the multivariate analysis of variance will often be a more informative approach than the univariate approach" (1971, p. 282-3), subsequent analysis hinged on multivariate analytical techniques.

Multivariate Analysis of Variance

The multivariate analysis carries with it the assumption that the levels of stimulus type must be considered distinct (though) correlated variables. The procedure is, essentially, a test of differences among profiles of means, where the profiles are composed of the means of responses to stimulus categories. Thus, for instance, the multivariate test of a subject effect will be a test of differences in the subjects' mean responses to the categories, and in that sense, is related to the univariate subject by type interaction. Statistical considerations apart, the relation between univariate interactions and multivariate main effects is a distinct advantage, since the primary theoretical questions involve the effect of manipulation of the stimuli on the differential response to stimulus categories.

Two analyses were performed on the standardized data, one involving all stimulus categories and another including only categories of ambiguity. At this point the issue of a fixed effects or a random effects interpretation of subject responses became pertinent. The use of multivariate procedures had been justified by the peculiarities of the variance-covariance matrices, the nature of these peculiarities were primarily a highly significant correlation between responses to ambiguous and nonambiguous stimuli. If this were interpreted as representing the influence of the form of response required of the subject (i.e. 'yes' or 'no', depending on presence or absence of ambiguity), then a strong case would exist for arguing that the behaviour was experimentally influenced and that subjects might be treated as a fixed effect.

The results of the analysis on all stimulus categories revealed a significant list effect, $\underline{F}$ (8, 414) = 1.97, $\underline{p} <$ .05, and a significant position by list interaction, $\underline{F}$ (8, 414) = 2.08, $\underline{p} < .05$, under the fixed effects model. No other effects were significant, and under the assumption that subjects were a random effect no significant effects were found at all (at the .05 level). The results of the analysis on the categories of ambiguity revealed no significant effects at the .05 level, under either model. The interpretation of these results must be given in terms of the profiles of means across stimulus categories. Thus, one conclusion would be that among the categories of ambiguity there is nc significant variation from the profile of grand means for that category. The inclusion of the nonambiguous category, however, reveals significant differences between profiles across levels of lists by position. At first sight, this might bear out the implication of a peculiar effect on response associated with the ambiguous/nonambiguous dichotomy.

Further interpretation of these results requires knowledge of the profiles. Prior to an examination of the means, however, it is revealing to consider the results of independent analyses of variance on each individual stimulus category (see Appendix C for tables). For the category of lexical ambiguity, only the grand mean was found to be a significant effect, under the fixed effects model, $\underline{F}$ (1, 210) = 5.29, $\underline{p} < .025$. For the surface ambiguity category, only list was found to be a significant effect, $\underline{F}$ (2, 210) = 2.69, $\underline{p} < .10$, under the fixed effects model. No significant effects were found for the underlying category, at the .10 level. While for nonambiguous stimuli, significant effects, under the fixed effects model, were revealed for the grand mean, $\underline{F}$ (1, 210) = 5.93, $\underline{p} < .025$, for list, $\underline{F}$ (2, 210) = 3.63, $\underline{p} < .05$, and for the list by position interaction, $\underline{F}$ (2, 210) = 4.55, $\underline{p} < .025$.

The results above have very strong implications. First, the tests on the grand means were tests on the difference between the mean and zero; the analysis was performed on standardized data, in which case the mean for each subject across categories was zero. As a consequence, the failure to find significant grand mean effects for both the surface and underlying ambiguity categories has the implication that the response times to these categories were generally equal to the subjects' mean response time. Furthermore, the absence of any significant effects in the MANOVA on ambiguous categories implies that the list effect under the surface category may be neglected and that therefore no evidence exists of a differential response between the surface and underlying categories of ambiguity. Similarly, the significant grand mean for lexical ambiguity implies a differential response between that category and the two categories of structural ambiguity, which, since the mean is negative, may be interpreted as indicating generally faster response than to structural ambiguity.

The interpretation of the list by position interaction in the MANOVA on all stimulus categories is a problem of some concern; interactions in MANOVA are notoriously difficult to interpret. The results of the ambiguity MANOVA and of the univariate analyses show that the significance of the interaction is contingent upon the presence of the nonambiguous category. The pattern of significant effects within the complete MANOVA and the ANOVA for nonambiguous stimuli is identical. Thus, whether or not the correlation between ambiguity and nonambiguity plays an important role in the significance of this effect, the role of the pattern of response to nonambiguous stimuli is crucial. However, the relative order of clauses is meaningless with respect to nonambiguous stimuli. It might be possible that the relative order of occurrence of main and subordinate clauses would be involved in a semantic effect that would imply interpretation of the position by list interaction as an expanded list (i.e., semantic) effect. However, the absence of comparable effects among the ambiguous stimuli suggests that the effect has to do with speed of determining that no ambiguity is present, which introduces an element of subjective judgement and raises the possibility of an effect which is essentially a subject effect. As a final note, it should be stated that an a posteriori comparison of means (Tukey 'a') revealed that the significant interaction in the univariate nonambiguity analysis was associated with three homogeneous groups of means, at the .01 level; P2L1 (Position 2, List 1) and P2L3, with relatively high means, P1L1, P1L2, P1L3, with intermediate means, and P2L2 with a relatively low mean. If the semantic relation between main and subordinate clause were the primary explanation of this effect (rather than subjects' judgements), then only two homogeneous groups of means would have been anticipated.

Discriminant Function Analysis

Thus far the results reported have been based on data standardized by subject. Since it has been determined that the effects due to the list and position interaction can not exclude consideration of a subject contribution, it is possible to make comparisons across subjects neglecting the nesting of subjects within levels of list and position. Furthermore, since the formal distinction between the surface and underlying categories of ambiguity was known to be unjustified (Prideaux, 1972), the absence of evidence of a differential response between them provides motivation for collapsing the two categories into one category of structural ambiguity.

A discriminant function analysis was formulated with the following structure: each subject was treated as a distinct source of variation in response, and responses were grouped under the categories lexical ambiguity, structural ambiguity, and nonambiguous sentences. The aim of the analysis was to find the linear function of the subjects' responses that would maximize the distance between the centroids of the three categories. Given the multivariate distribution, two significant functions (or dimensions) would be required to represent a joint differential response to all three categories. Thus, the results of the discriminant function analysis would be profiles of centroids which would represent the characteristic response of a population of subjects, and would indicate which groups were responded to differentially.

The outline given above was couched in general terms because it was quickly found that several subanalyses were required. An attempt was made to determine the response profile that would characterize the complete set of subjects; however no significant discrimination was found.

This was attributed to the presence of a significant proportion of subjects who did not in fact respond differentially to the categories. A series of one-way ANOVA were performed, one per subject; and it was found that only eight of the total of 42 subjects had significant category effect F-ratios at the .05 level. Using a very liberal criterion, that a subject's F-ratio must be greater than 1.0, a subset of 20 subjects (termed 'discriminators') were selected for further analysis.

The calculation of the discriminant function for the discriminators revealed only a single significant dimension, at the .05 level. In contradiction with the results of the MANOVA, differential response was indicated between structural ambiguity and the pair lexical ambiguity and nonambiguous. Since the mean response time for nonambiguous sentences was greater than any of the other categories, this result appears to have the implication that by weeding nondiscriminating subjects out, responses to nonambiguity become comparable to that of lexical ambiguity. In order to examine this question, the subject standard score means for each category were multiplied by the appropriate discriminant function weight and the resulting scores graphed (see Figure 2). It is immediately apparent that the centroid profile is not representative of overall subject response and that the significant dimension of discrimination is likely due to a small number of comparable outliers.

Having found that a satisfactorily representative response profile was not obtainable from either the full set of subjects or the full set of discriminators, the discriminators were further subdivided. The standardized subject profiles were submitted to a hierarchical clustering routine, that utilized Ward's adaption of the Veldman algorithm (DERS, Test 10). The optimum clustering appeared to be that which involved subdivision into four groups. These groups' responses were analyzed, and the distribution of weighted subject means about centroids were graphed as above.

The largest group of subjects (Group 1), seven in all and graphed in Figure 3, had a single significant dimension of discrimination, at the .1 level. The relatively low level of significance might be explained by the inclusion of two subjects whose responses appear to be marginally deviant from the other subjects in the cluster. The pattern of response implied by the centroid profile indicated a dichotomous response with relatively rapid identification of lexical ambiguities with a slower response to all other stimuli.

The second largest group (Group 2), comprising six subjects and graphed in Figure 4, had a single significant dimension at the .1 level. This group appeared to have a generally rapid response to ambiguity with a slower response to nonambiguous stimuli.

The third group (Group 3), containing five subjects and graphed in Figure 5, had a single significant dimension at the .03 level. This group appears to have reversed the strategy of the second group, having a relatively slow response to categories of ambiguity and a quicker identification of nonambiguous stimuli.

The final group (Group 4), graphed in Figure 6, contains only two subjects. The justification for attributing group status to these subjects was that addition to any of the other groups would have involved an unacceptable increase in the within-group error calculated by the hierarchical clustering routine. The pattern of discrimination, a single dimension significant at the .002 level, appears to be similar to that of Group 1, involving differential response between lexical and structural ambiguity. However, in this case, the use of the discriminant function to reclassify the responses would have resulted in a proportion of the response to nonambiguous stimuli being misclassified as either of the lexical type or of the structural type. In the case group , responses to nonambiguous stimuli and to structural ambiguities would have been grouped together.

These results may be interpreted as indicating three general patterns of response, each one being dichotomous. Groups 2 and 3 that responded differentially to the ambiguous/nonambiguous dichotomy represent the correlation observed in the earlier stages of analysis. It does not seem unreasonable to interpret this correlation in terms of the consequences of bias for or against detection of ambiguuty, given sensitivity to the ambiguous/nonambiguous dichotmy. The other two groups appear to have been relatively insensitive to the ambiguous/nonambiguous dichotomy, Nevertheless their respective treatment of nonambiguous stimuli is crucial to a characterization of their responses. Group 1 responded uniformly to both structural ambiguity and nonambiguous stimuli, which might be interpreted as a lexical ambiguity/'other' dichctomy. The tendency for Group 4 to respond to some nonambiguous stimuli in a manner similar to the structural category, suggests that these two subjects were the only ones genuinely responding differentially to the grammatical dichotomy between the types of ambiguity.

Since the original experimental design nested subjects under levels of list and position, it was possible to identify the list with which these subjects had been associated and to attempt to relate the patterns of response to the list by position interaction that had been found to be significant in the overall MANOVA. The subjects were identified as being drawn from one of the high, low or intermediate groups of list by position means within the nonambiguous category. It was then assumed that the hierarchical clustering routine could be construed as sampling without replacement from these three groups. On this basis, using the hypergeometric sampling distribution, it was possible to calculate exact probabilities that subjects who were clustered together also were associated with the same list by position group. It was found that the two groups of subjects who did not appear to respond to the ambiguity/nonambiguity dichotomy had a relatively high probability of representing a random sample from the groups of list by position means (the probabilities were .1918 for Group 1 and .1474 for Group 4). However, Group 2 was drawn from the high range list by position group (5 out of 6 subjects), and the probability of this occurring at random was .0116. Similarly, Group 3 was largely drawn from the low range list by position group (3 out of 5 subjects, with 2 from the mid range), and the exact probability of this sample was .0072. On this basis, the extreme (either high or low) responses to nonambiguous stimuli seem to be associated with subjects who were responding to the ambiguous/nonambiguous dichotomy. This result would support the view that sensitivity to this dichotomy is associated with a task bias, since the responses of these subjects to nonambiguous stimuli were extreme by comparison to the subjects sensitive to the other dichotomies. It is noteworthy that the potential for such a task bias might have been introduced by instructing the subjects to give special care in their responses, since there were more ambiguous stimuli than nonambiguous with the implication that guessing would generally result in error.

Analysis of Covariates

The Covariates

Given substantial information about type effects on reaction time, the goals of the multiple regression analysis were to consider jointly other sources of variation and their interactions with the type effects. The other sources of variation to be considered were broadly related to either subject differences or error in response. The justification for considering sources of subject difference was the heterogeneity of subject variances in the response time analysis. Moreover, even after standardization the remaining effects were seen to be associated with differences between subjects. Thus, subject effects were known to be a systematic source of variation in response time.

The use of covariates which would reflect the error in response was theoretically motivated. First, it was of interest to see whether the results reported in Prideaux and Baker (1976) concerning the distribution of errors with respect to stimulus type had been replicated in this experiment. Additionally, the implications of that experiment taken in the context of the implications of a speed/accuracy trade-off as outlined in CRT theory, make the investigation of the relationship between error and time crucial to a characterization of the processes at work.

The following are descriptions of the independent variables whose covariance with reaction time was to be investigated:

a) Sequence (SEQU): Since each subject was presented with a uniquely randomized stimulus order, the sequence effect could be examined independently of the nature of the stimuli. If significant, this effect would represent changes in subject response associated with learning or practice.

b) Ambiguous/nonambiguous dichotomy (AMB): The dichotomous effect detected in the previous analysis.

c) Ambiguity type dichotomy (TYPE): as above.

d) List by Position by AMB interaction (LAMB): as with (b) and (c) above, this effect represents the results of the previous analysis and is known to be significant. The inclusion of these variables is intended to prevent spuriously redundant predictions and to ensure that the information gained by this analysis is new information.

e) Stimulus categorization (ERROR): This variable represents the correctness or incorrectness of the subjects' identification of a stimulus as either ambiguous or nonambiguous.

f) Semantic interpretation (PARA): In this instance, the correctness of the paraphrase or semantic distinction that the subject described was represented.

g) Semantic bias (BIAS): This variable represented the results of the preferential judgement task outlined as experiment 2 in the Methods chapter. The percentages that were estimated from that task were normalized by the arcsin transformation, and all nonambiguous stimuli were treated as having 100% bias.

h) Sex (SEX): Given the flexibility of multiple regression analysis it was now possible to determine whether some of the intersubject difference could be attributed to a sex difference.

i) Subject bias (LKRT): This variable was the most complicated of the covariates, involving a non-linear transformation of relative proportions of error, but seemd to be justified by the suggestions of a subject bias effect in the previous analysis. The extreme response times to nonambiguity appeared to represent a tendency to opt for identification of one response category in preference to another, independently of the properties of the stimulus involved. Such an effect can be modelled in terms of the principles of Signal Detection Theory, in which the probability of a correct response to a signal can only be meaningfully evaluated relative to the probability of the same response in the absence of a signal. In the present case, this may be interpreted as the probability of identifying ambiguity correctly relative to the probability of claiming to have detected an ambiguity in a nonambiguous stimulus. Following the techniques of Signal Detection Theory, likelihood ratios were calculated for each of the categories of ambiguity, by taking the ratio of the proportion of correct ambiguity responses to the proportion of incorrect nonambiguity responses for each subject and category (in those cases in which the proportion was zero, the value 1/2n was used). In order to adjust for the nonlinearity of the likelihood ratios, the natural logarithms of the ratios were used. In order to make the most efficient use of this data, a repeated measures anova was carried out involving ambiguity category and sex factors. The repeated measures assumptions of homogeneity and compound symmetry were satisfied at the .25 level. Only the ambiguity category effect was significant, at the .10 level, and a 'Tukey a' procedure revealed a significant difference between lexical and underlying ambiguities, with the implication that lexical ambiguity was less subject to this bias than the stimuli categorized as underlying. Since the surface category was intermediate between lexical and underlying, it was assumed that the previous justification for disregarding the putative distinction between surface and underlying ambiguity could also be applied in this case. Thus, as data, each subject's bias was divided into a log likelihood ratio for lexical and structural ambiguity, and an additional log likelihood ratio was calculated to represent the deviation of incorrect nonambiguous responses from random. Sufficient variety had thus been introduced to avoid spurious correlation between the log likelihood ratios and differences between simple subject mean response times.

In addition to the variables described above, the interaction between all of the independent variables and the stimulus type effects (AMB and TYPE) were incorporated into the analysis, with the exception only of interactions with

SEX and LKRT. There was no theoretical basis for including or interpreting an interaction with SEX, and in the case of LKRT, it was judged that as a consequence of the preliminary analysis, it already constituted an interaction effect.

Multiple Regression

The overall regression results were significant, *F* (15, 992) = 11.998, $p < .001$; however, the proportion of variance of the log response times accounted for by the regression was relatively small (modified $R^2 = .14076$). The individual variables that were found to be significant predictors of log response time were: LKRT, $p < .01$; SEX, $p < .001$; ERROR, $p < .001$; LAMB, $p < .025$; PARA, $p < .001$; AMB, $p < .10$; SEQU by TYPE, $p < .10$; and ERROR by AMB, $p < .001$. A more detailed account of the results may be found in the statistical appendix (C). A number of ancillary analyses were performed in order to ensure that none of the effects cited above were spurious. These analyses, involving an investigation of the effect of interaction terms and an investigation of the interrelations among independent variables, are detailed in Appendix C. No reason was found to doubt the validity of these results for the given data.

The interpretation of the significant effects in terms of the associated regression coefficients were as follows:

a) LKRT: The existence of a positive regression coefficient indicates that as likelihood of correct response increases the response time also increases. The effect for a given subject will in general be different depending on the category of stimulus involved, since this variable represents both subject specific bias and a differential ambiguity type effect.

b) SEX: This effect may be interpreted as indicating a generally longer mean response time for females than for males. Ancillary analysis indicated redundancy between SEX and LKRT, which may indicate a sex effect in error scores as well, or possibly might suggest that females were generally more responsive subjects.

c) ERROR: The appropriate interpretation of this effect was that the mean response time for incorrect responses was generally longer than that for correct responses. This contradicts the prediction of a speed/accuracy trade-off.

d) PARA: As with ERROR, the interpretation of this effect associates longer response times with incorrect responses. However, the compatibility of the results for ERROR and PARA may be taken as evidence that the processes involved in identifying the presence of ambiguity have sufficient in common with the processes of ambiguity interpretation to be used as an estimate of interpretation facility. The regression coefficients of ERROR and PARA are nearly identical (-0.15 and -0.12, respectively, for standardized scores) and the fact that they are independent predictors may be interpreted as reflecting sensitivity to different levels of a common process.

e) ERROR by AMB: The interpretation of this effect required ancillary analysis which revealed that the effect described for ERROR and PARA was differential depending on whether a stimulus was ambiguous or nonambiguous. After the removal of simple ERROR and PARA effects, the relation between ERROR and response times to nonambiguous stimuli was not significant; however, incorrect responses to ambiguous stimuli were still associated with longer mean response times than correct responses. The differential ambiguity type effect represented by LKRT must be borne in mind when the relevance of this effect is evaluated, ancillary analysis indicated a relationship between these two independent variables.

The finding that BIAS was not a significant effect was of theoretical interest, particularly since it had been significantly correlated with response time. The fact that the relationship had been partialled out suggested that another variable or variables represented the effects to be accounted for by BIAS more directly than BIAS did. The variables involved could be identified by calculating the partial correlations between BIAS and the other independent variables, this was done neglecting interactions with BIAS. One of the variables identified as being involved was AMB, the reason for this involvement was most likely that nonambiguous stimuli were coded with 100% bias. More interesting were the contributions of TYPE ($p < .001$) and PARA ($p < .06$). Based on the sign of the partial correlation, the interpretation for TYPE would be that a higher bias tended to be associated with lexical ambiguities, or alternatively subjects were more variable in their preference judgements with structural ambiguities. The interpretation of PARA would be that excluding the covariance associated with all other independent variables there was a tendency for subjects to correctly paraphrase those stimuli associated with high bias. These conclusions are not inconsistent with the assumptions about the information that the bias measure was supposed to represent. However, it appears that the variable is wholely redundant and the information is more directly represented by other variables. It is noteworthy that the relationship between TYPE and BIAS is the only effect that appears to be consistent with the results of the M&B study, and as such suggests that their simple correlation between bias and response time was a consequence of a mutual relationship with ambiguity category.

A final note on this analysis, an analysis of the residuals of the regression indicated a significant Durbin-Watson statistic (Draper and Smith, 1966). In general, this could come about in one of three ways: (a) there could be an undetected nonlinear relation between the dependent variable and at least one of the independent variables; (b) there is a genuine stochastic dependency between successive response times; or, (c) that a significant source of variation (another variable) has not been identified.

CHAPTER FOUR

DISCUSSION

Sentence Processing

Three aspects of M&B's characterization of sentence processing can be evaluated in light of the results of this study. They are the assumptions (1) that the taxonomy of ambiguities corresponded to stages of processing, (2) that there exists a unique 'psycholcgical order' of the processing of stimulus material, and (3) that processing was carried out as a rule governed function of grammatical analysis. The results of the analysis of data from the present study appear to conflict with each of these assumptions, and the nature of each conflict will be examined in the following sections.

Ambiguity Type and Stages of Processing

There are two aspects to the comparison of responses to different categories of stimulus; the first is the direct comparison of means under stimulus categories, and second is the comparison of variability among subjects within categories. The results of the MANOVA on standardized data among ambiguous types appear to justify the acceptance of the overall mean standard scores as the basic response profile, after differences in base response times and degrees of variability between subjects were accounted for.

On this basis, no significant difference was found between the 'underlying' and the 'surface' categories of ambiguity. M&B claimed that these two categories represent distinct stages of processing, whereas in this study their respective means were both found to be equal to the standardized subject mean response time. Since it was noted in the outset that no formal argument could show the necessity for the distinction between these two categories of ambiguity, it was assumed that they could be collapsed intc a general category of 'structural' ambiguity. At this point the stage-of-processing explanation for the differential response between the categories breaks down, since the two categories could equally as well be said to represent differences in the kind of information processed as the stage of processing.

Since the overall mean of 'structural' ambiguity is related to the subject mean response time, the question of whether this category might represent an individual stage of processing can be settled by determining whether subjects' standardized means are consistent estimates of the overall mean. If subjects did not respond consistently to the stimulus category, it could not be claimed that they were processing at a characteristic stage. The average consistency of response is related to the significance of the intra-class correlation coefficient (Haggard, 1958), which is related to the F-ratio in the following manner: (a) if responses are consistent, the variances associated with an effect will be partitioned so that MS-within will be smaller than MS-between, (b) if responses are random, MS-within will approximately equal MS-between, and (c) if responses are inconsistent, MS-within will be greater than MS-between. The latter case could be considered as a test on the F-ratio to determine whether it is significantly less than 1.0. Unlike more common correlations, the intra-class correlation has a lower bound which is determined by the number of within-class observations (calculated as $-1/(k-1)$ where $k-1$ is the within-class degrees of freedom). For the present case, the lower bound was related to the number of subjects and was calculated to be -0.0277. The intra-class correlations between subjects were found to be significant for each of the ambiguity categories: lexical, $r = 0.0993$, $p < .025$; surface, $r = -0.0243$, $p < .001$; underlying, $r = -0.0240$, $p < .001$. In the case of both forms of structural ambiguity the correlation coefficient is near its lower bound; similarly, a test on within-subject consistency revealed significant inconsistency under the structural ambiguities (at the .05 level) and randomness under the lexical category. Thus, despite standardization of the data, the variability both between and within subjects can not be attributed to random variations in response to structural ambiguity. Therefore, the grand mean of structural ambiguity can not be a parametric estimate of standard response time, as it would be were it associated with a distinct stage of processing. It seems necessary to conclude that the crucial distinction represented by the ambiguity taxonomy is the distinction between types of information (e.g. lexical or structural relations), and that the processing of structural information has inconsistent subjective response time effects associated with it.

Order of Processing

The theoretical position adopted by M&B implies that the notion of processing difficulty is directly related to the order of processing as inferred from the response times. As difficulty was assumed to be typologically invariant, so the order of processing was assumed to be unique. However, since the M&B hypotheses were directed exclusively towards the distinctions between ambiguity types, it does not seem justified to reject their claim as a consequence of the finding that strategies differ with respect to the involvement of nonambiguous stimuli. Thus, the two groups of subjects who were found to respond to the ambiguous/nonambiguous dichotomy (Groups 2 and 3) are irrelevant to this issue.

The two remaining groups (Groups 1 and 4) were characterized as responding to a lexical/other and a lexical/structural dichotomy, on the basis of their treatment of nonambiguous stimuli. However, it may not necessarily be the case that the involvement of nonambiguous stimuli is crucial to the characterization of the dichotomy. It is possible that both groups represent a strategy in which the lexical information is processed first, and in the cases that ambiguity is not found the structural information is processed as an alternative. This strategy would correspond to a word-level processing with subsequent phrase level processing, and would imply that group 4, who responded as fast to some nonambiguous stimuli as to lexical stimuli, would likely have more incorrect responses to nonambiguous stimuli than group 1. The latter prediction was borne out, although considerable risk is involved in attaching significance to the responses of only two subjects.

On the other hand, the regression analysis revealed an ambiguity type effect only in interaction with the sequence of presentation over the course of an experimental session. The interpretation of the regression coefficient is that the relative differences in response times between lexical and structural ambiguities increase as the experimental session progresses. This effect might be interpreted to mean that the word level followed by phrase level processing strategy is not inherent, but only acquired with practice in this type of task. However, M&B explain the response time discrimination effect in terms of difficulty, and when interpreted with respect to difficulty the regression effect implies that structural ambiguities become relatively more difficult as the experimental session progresses. This latter interpretation, though seemingly contradictory, might have an explanation. The sequence by ambiguity type interaction was found to be significant only if partialled with the nonsignificant sequence by ambiguous/nonambiguous interaction. This suppression effect (see Appendix C) may represent the involvement (i.e., interference) of the distribution of nonambiguous stimuli across trials in the formation of a strategy which leads to the differential response to types of ambiguity. This involvement might have the following character: subjects were made aware of the fact that more ambiguities would be presented to them than nonambiguous stimuli, thus in effect they were instructed to respond 'yes' in more than 50% of the trials; if near the end of a session, a subject did not feel that he had responded 'yes' a sufficient number of times, this could represent evidence to him that there was a greater a priori probability of the final stimuli being ambiguous and consequently he would look harder and spend more time. If in general it is more difficult for a subject to distinguish between nonambiguous and structural stimuli than between nonambiguous and lexical stimuli, then relative differences between ambiguity types increasing across trials would be predicted. If response times to lexical stimuli are assumed to remain constant and only those for structural stimuli increase, there would be an explanation for the inconsistency of subject responses within the structural category. Furthermore, the idea that structural and nonambiguous stimuli are relatively difficult to distinguish may be consistent with the relation between BIAS and TYPE. Structural ambiguities were found to have greater inter-subjective variability in preferred interpretation than lexical ambiguities. It would be reasonable to assume that if variations in interpretation of nonambiguities exist, these would also be inter-subjective effects.

The explanation of differential response to stimulus categories in terms of intrinsic ordering of processes is not borne out in these results. From the preceding section, it is known that such ordering would be unlikely to be an ordering of distinct stages of processing. The alternative of ordered processes involving distinct types of information can not be given an unqualified interpretation. At each point in the analysis, it emerged that influences from all stimulus categories interferred with response to any given stimulus. Whether this relativity of response is an experimental artifact or inherent to the process is a question that requires more information about the nature of processing.

Processing: Deterministic or Probabilistic

The assumption that sentence processing is deterministic in character underlies the whole 'psychological order' model as proposed by M&B, as well as being implicit in the use of such terms as 'meaning perception'. To be taken seriously, the assumption would require that the processing of a given stimulus be independent of the decisions reached for other stimuli, be free of subjective effects, and that errors be due to random influences. It has become clear in the preceding sections that the processing of structural ambiguities in particular is not independent of the processing of nonambiguous stimuli. A similar dependent response to lexical ambiguities is represented by the subjects who responded to the ambiguous/nonambiguous dichotomy. The behaviour of these groups of subjects might be characterized in terms of a search strategy; if a subject happened to have a subjective bias against detecting ambiguity, he might delay decisions given ambiguous stimuli (and vice versa, given a bias _for_ detection of ambiguity). The interpretation that processing is in part governed by a preferred response is consistent with the finding that relatively long response times to ambiguous stimuli are significantly correlated with incorrect responses (as represented by the significant ERROR, PARA, and EAMB regression effects). These effects have the character of a search in the sense that processing may in part involve finding evidence to justify a predetermined preferred response. It may, however, be a consequence of the forced-choice character of the task and that quite a different picture would emerge if subjects' responses were in the form of confidence ratings of their judgement of the likelihood of a given stimulus being ambiguous.

Since all of the subjects for whom there is evidence of a differential response were found to have responded dichotomously, the potential for a relationship between the nature of the processing and the response alternatives arises. It is possible that these dichotomous strategies are exclusively dichotomous because the response is dichotomous, which suggests that subjects have generally processed only sufficient information to draw the distinction required of them. The suspicion that those subjects ('discriminators') who were included in the discriminant function analysis represent a distinct class of subjects, and that the remaining subjects may have misunderstood the task is contradicted by an examination of each subject's intra-class correlation. On this basis (at the .05 level), it was found that nine subjects responded consistently, 25 responded randomly, and the remaining eight responded inconsistently. Such a distribution might imply that subjects were drawn from a random population (as regards response consistency). When the specific kind of strategy is considered and the correlation between subjects is taken into account, then the subjects fall into distinct groups. On this basis, it may be legitimate to conclude that the 'discriminators' do not form a special case (they are sampled from a random population) and therefore the propensity to respond dichotomously is a population characteristic not a property of the distinct strategies associated with sub-groups. Thus, it would be of interest to carry out further experimentation in order to determine whether increases in the number of response alternatives would result in greater complexity of differential response. An increase in the number of response alternatives might be achieved by requiring subjects, after appropriate training, to either name the stimulus type (i.e., lexical, structural, or nonambiguous) or to rank their responses in terms of confidence about the presence of ambiguity (perhaps with scales of varying intervals).

The direct interpretation of the character of processing based on the relation between errors and response time may have only limited validity. The comparison of these two sets of observations is essentially a comparison of stimulus associated effects, while processing is a subject characteristic. For this reason an apparent contradiction in the regression results is tolerable. The error/response time comparison suggested that incorrect responses were associated with long response times, while the interpretation of the LKRT variable was that correct responses relative to the responses to other categories were generally associated with subjects whose response times were relatively long. This latter finding is of particular interest in that it is consistent with the Statistical Decision-making model proposed by CRT theory. The Statistical Decision-making (SD) model predicts that if response time interval represents the size of an information sample that a subject processes, correct responses will generally increase as a function of the sample size. The SD model predicts a direct relation between errors and response times for a given subject who is free to set his own criteria for decision-making. The criteria are represented in terms of likelihood ratios, where the magnitude of the critical values (criteria) of the likelihood ratios will determine the average sample size required by a given subject to reach a correct decision. Thus, the LKRT variable will represent an estimate of the criteria that a subject may have used in decision-making, and it is the appropriate variable to test the hypothesis that a group of subjects with differing criteria responded in a manner consistent with the SD model.

The assumptions above require some qualification, since the analysis of variance on likelihood ratios revealed a significant difference between the lexical and underlying categories (as consistent with Prideaux and Baker's (1976) analysis of response errors). The inclusion of independent values associated with lexical and structural types for each subject suggests that these were independent criteria and thus, that independent SD tests were carried out on the different stimulus categories. This would, however, conflict with the implication that may be drawn from the general dichctomous responses, that subjects for the most part considered only two alternatives in examining the stimuli. The investigation of relationships among independent regression variables revealed significant mutual enhancement of LKRT and LAMB (which was the variable representing subject group effect). This relationship raises the possibility that underlying the likelihood ratio analysis of variance effect was a subject by type interaction which corresponded to the subject groupings represented in the results of the discriminant function analysis. Due to the paucity of data, such a relationship cannot be tested for these subjects responses; however, it is noteworthy that in the two groups that responded to the ambiguity types, the group of two (that seemed to respond inconsistently to nonambiguous stimuli) had low LKRT scores for nonambiguous stimuli, while the other group had high scores; and in general, subjects appeared to have a lower likelihood of correctly identifying nonambiguous stimuli, than ambiguous. The latter effect may be associated with the relationship found between LKRT and EAMB, and may in part be explained by the possibility of a relationship between a subject's response bias and the (relative magnitudes of the) criteria imputed to the (response alternatives of the) SD process. It should be borne in mind that the SD model has much in common with a general search strategy, with the difference that it allows termination of a search with an uncertain conclusion.

It would seem that much of the interpretation of the results of this study points towards a non-deterministic SD model of processing. This being the case, subject behaviour might be better characterized as signal detection than as computation based on grammatical rules. If a signal detection paradigm were the most appropriate paradigm for this task, it would follow that the results of this experiment would be inadequate for a comprehensive description of sentence processing in this context. Among the weakest features of the experimental design would be the lack of control over subject expectancies about the proportion of ambiguous to nonambiguous stimuli. It would be of interest to see whether inter-subject variability could be reduced by explicitly stating the proportion. Similarly, it would be of interest to know what effect it would have if subjects were deliberately misled about the proportion.

It is possible, on the basis of these results, to propose a processing model which is considerably different from that postulated by M&B. In the first instance, the notion of 'difficulty' may be introduced, but only as a relative term; that is, two stimulus categories may be more difficult to distinguish than another pairing of stimulus categories. Thus, difficulty would not completely explain differential temporal response. In order to explain differential response, at the least information about accuracy of response, response bias or preference, and subject characterization of alternatives must be provided. The latter consideration arises from the difficulty of expressing the differences that might have led to the different dichotomies to which subjects responded. In terms of the SD model, the explanation of the difference between the groups that responded to the ambiguous/nonambiguous dichotomy as a differing response bias is perfectly acceptable. An explanation of the responses of the other groups might include a tendency to regard ambiguity as being primarily lexical, which would be consistent with seemingly greater accuracy in response to lexical ambiguities for these groups (which was not the case for the ambiguity/nonambiguous dichotomy). Thus, a differential temporal response is a function of context sensitive or experientially based biases, rather than uniquely ordered stages of processing; and the fundamental process involved would be informed guessing.

Relation to Issues in the Literature

Evidence has been described for effects of a totally different character from the model developed by M&B. However, two more general issues given frequent comment in the literature need to be examined. These are the status of token effects, as represented by BIAS, and the issue of the relative complexity of processing ambiguous over nonambiguous sentences.

The analysis revealed no systematic effects whatsoever that could be attributed to tokens. This was found both for variations in presentation lists and for the semantic bias measures gathered. The analysis seemed to suggest that insofar as variations in semantic bias were regular, they reflected the effect of the type of information on which they were based and thus, the ambiguity type. The finding that BIAS was not a significant independent predictor of response time was of particular interest in light of the manner in which the measures were obtained. The semantic bias measures were gathered in a manner which assumed that these token-like effects would be inherent properties of each token. The failure of these estimates to predict response time replicates the finding of Foss, Bever, and Silver (1968) that pre-test evaluation of semantic bias did not yield estimates that correlated with response time. They found, on the other hand, that estimates of semantic bias gathered during the test were related to response time. This would be consistent with the context sensitive behaviour observed in the present study, and would lend support to the impression that the present results are not aberrant.

Several of the models proposed in the literature raise the question of whether ambiguous sentences are inherently more or equally complicated to process than nonambiguous sentences, and in general it is understood that complexity would be reflected in response time. Given these preconditions, the answer that would be inferred from the present results would be that depending on the subject, nonambiguous sentences can be more, less, or equally as complicated to process as one or all of the types of ambiguity. Such an answer might well be taken to suggest that the question was ill-conceived. The models have for the most part drawn heavily on the structure of transformational grammar. However, from one point of view, that structure deals only with well-formedness constraints and not the relation between meanings and representations. It is a well accepted principle of interpretive semantics that the mappings between representations and meanings are governed by both the structure of the representations and presuppositions. Yet, by the nature of the experiments performed, the contextual guidelines for interpretation are missing. Though the effect that this might have on response times is unpredictable, it might be reasonable to interpret the complexity of the present results as being in itself support for the notion of an interpretive semantics. If in particular response bias is related to the notion of presuppositions, then the present results might directly reflect the interaction between the two components of interpretive semantics. In this sense, it might be suggested that there would be much gained from constructing psychological models which conform to the spirit rather than the form of grammatical theories.

Difficulty and CRT Theory

The explanation that has been proposed for the differential response between ambiguity types in this thesis incorporates a notion of difficulty which is not absolute processing difficulty, but is instead the relative difficulty a subject would have in distinguishing one or other of the ambiguity types from the nonambiguous control. Furthermore, it is proposed that the differential response effect that arises from this relative difficulty will itself only be evident as a consequence of a response bias varying across trials. It is noteworthy that the responses of subjects for whom there is no evidence of discrimination would from this perspective be regarded as unbiased responses.

This notion of difficulty bears much in common with the general task difficulty developed in CRT theory. The predictions derived from CRT theory would suggest that difficulty and therefore mean response time is directly related to the number of stimulus categories; implying that in general, it is more difficult to distinguish between three categories than it is between two. Furthermore, it is possible, given certain assumptions about the nature of processing, to derive predictions about whether increases in the number of stimulus categories would lead to increases in response time as a linear or a logarithmic function of the number of stimulus categories. Since it would be possible to train subjects to respond to the three classes of stimuli, it would be of interest to discover what sort of increase in difficulty this would entail.

However, the general question of difficulty, as it arises in CRT theory, does not only bear on the question of the relative magnitudes of response times. There are in addition predictions concerning the form of random distribution that response times can take:

> ... it is a relatively common observation that as certain decision situations are made more and more simple, the observed latency is better and better approximated by an exponential distribution slightly displaced from the origin. (Christie & Luce, 1963, p. 25)

In order to examine the present data in this light, histograms were constructed representing the number of times that subjects failed to make a decision within successive two second intervals (see Figure 7). Using a discrete counterpart of the exponential distribution , it was observed that a relatively adequate fit could be achieved. It is noted by Christie and Luce that the deviation of the empirical from the theoretical curve will in part depend on whether processing could involve serial or parallel processing. Such relationships would be useful in comparison with results of further experiments to settle the question of order in sentence processing. Nevertheless, the relatively slight deviation from the theoretical curve may be accepted as evidence that the number of elementary decisions involved in reaching the stimulus categorization is small. This would be consistent with the interpretation of the present results in terms of the SD model or more generally in terms of a search strategy. However, the crucial test of more specific properties of the processes would require examination of the effects of increasing the number of stimulus categories to which a subject would have to explcitly respond.

Processing Dynamics and Linguistic Theory

If the results of this experiment are not to be thought completely aberrant, then one general implication emerges: the processes which underlie sentence comprehension are highly flexible, and may even be considered to be adaptive to a speaker's information requirements. The dynamics of sentence processing might be characterized as being susceptible to constraints associated with both expectancies and the use to which the information will be put. In this sense, it may even be justifiable to characterize standard sentence processing as commonly involving the choice among multiple potential interpretations, a characterization which might suggest that sentences regarded as ambiguous are more nearly normal than those sentences that might unfailingly be identified as nonambiguous .

It is encouraging that the results of this experiment appear to be subject to the same sorts of considerations, both in terms of the psychological constraints that condition responses and in terms of statistical description, that are dealt with in more general psychological information processing theories (as for instance Stimulus Sampling Theory (Neimark & Estes, 1967) and CRT theory). This may carry with it the implication that sentence processing is governed by general psychological principles and not principles that are peculiar only to language. But at the same time, it would have implications for linguistic theory. If, as a case in point, the SD model of processing were to be taken seriously as a model of sentence processing, it would follow that the stimulus samples on which a subject based his decisions would have to be samples of overt stimulus characteristics. That is to say, that the subject must be sampling the information represented by the surface structure characteristics of the sentence. To claim that a subject did more than that would require that specialized language processing faculties be invoked and would violate the character of the SD process. In this sense, the most general results of this study lend support to the notion that the true meeting ground of linguistics and psychology is at the surface structure level.

CHAPTER FIVE

SUMMARY AND CONCLUSION

The goal set for this study was the replication and extention of an experiment dealing with the behavioural impact of sentence ambiguity as reported in M&B. The results of that experiment had been interpreted as supporting the general hypothesis that the processes leading to sentence comprehension are rûle governed in a manner consistent with grammatical theory and that consequently the difficulty of comprehending the alternative interpretations of ambiguous sentences is directly reflected in the amount of time required to identify the alternatives. Much of the justification for M&B's conclusions rested on the validity of postulating the existence of a distinct category of ambiguity, in which deep structure information is necessary for the resolution of the ambiguity. An examination of the issues and arguments concerning this category suggested that from a formal standpoint the category was not well-founded, and that M&B's study failed to provide empirical evidence of the category's putatively distinct, behavioural impact.

In this study, M&B's experimental design was modified to include nonambiguous control sentences and a control for the potential effects of semantic material. Nevertheless, the major source of variation revealed was intersubject variability. Examination of the data revealed evidence that intersubject variability in response time was associated with variations in the accuracy of response, and with response biases that might be associated with differing strategies of response. Thus, in the main, M&B's assumption of deterministic processes (i.e., rule governed) was contradicted. The other assumption that difficulty of processing could be the conceptual link between the grammatical complexity of ambiguity categories and observed response times was contradicted by the finding that there was no empirical evidence to support the distinction between putative types of structural ambiguity (i.e., between 'surface' and 'underlying' ambiguity). Furthermore, the analysis appeared to favour an interpretation of difficulty in terms of structural ambiguity having greater similarity to nonambiguity than lexical ambiguity had, under certain response bias conditions.

At the outset, five areas of potential controversy were delineated. Two of these, including methodology and the relation between linguistic theory and the experimental design, were concerned with a critical examination of the M&B study and the present study has shown the criticisms to have empirical justification. The remaining areas, including problems related to naive speaker behaviour, reaction time as a dependent measure, and experimental control in psycholinguistic research, have a much wider scope than simply the study of the behavioural impact of ambiguity. Nevertheless the present study can serve to illustrate the significance of these issues.

The rejection of M&B's processing model, with its grammatically determined stages of processing, as a basis for the explanation of the present results adds considerable complexity to the theoretical problem. The pattern of responses observed in this study seems to imply that a characterization of sentence processing must include factors such as the strategy options available to a subject, the as yet unpredictable variations in accuracy of a subject's performance, and the nature of the response selection as well as the stimulus categorization task.

The general implications of the present results are that the form of behaviour under observation is a type of choice behaviour rather than the deterministic rule-governed behaviour, postulated by M&B. The strongest indication of this is that the stimulus effect associated with difficulty of discriminating between structural ambiguities and nonambiguous stimuli under certain conditions might be interpreted as implying that these two stimulus categories are the two most similar categories. Thus, difficulty represents the effect of interference in response between categories, rather than the effect of the properties or processing requirements of a given category.

The combination of choice behaviour with variable accuracy in subject performance has implications both for the relationship between psycholinguistics and grammatical theory and for the design of further research into ambiguity. In the first case, the fact that subjects make decisions with varying degrees of certainty suggests that the categorical distinctions made by grammarians are not necessarily distinct for subjects. The stimulus taxonomy with which this study dealt was inherently hierarchical; that is, the major distinction was between ambiguousness and nonambiguousness, while the lexical and structural distinction was nested within ambiguousness. However, the labels 'lexical' and 'structural' may represent the existence of a potential for ambiguity at some locus within a sentence, while the arguments of Ziff (1965, 1967), Gleason (1965), and Patel (1974) would suggest that a global evaluation of the sentence need not necessarily reveal ambiguity. Thus, the categorization of sentences need not be hierarchical, sentences may contain lexical or structural potential as within-sentence properties and independently may be ambiguous or nonambiguous as global sentence properties. For completeness, one might wish to speculate that the distinction between simple nonambiguous sentences and sentences that are in Quine's sense 'vague' could be the distinction between ambiguous and nonambiguous sentences that contain no (grammatically identifiable) potential. Such a cross-categorization could be the basis for an explanation of the dichotomous responses of subjects, as opposed to the potential three-way discrimination associated with the hierarchical nested categorization.

It is with reference to the categorization above that it may be most clearly seen that identification of ambiguity does not simply involve identification of grammatical properties within the sentence. Thus, the uncertainty of subjects' general performance seems less inexplicable. It is not necessary to infer from incorrect responses that subjects responded without understanding either the stimulus sentence or the experimental instructions. The existence of uncertainty raises the question of whether the forced choice dichotomous response was appropriate. If the categorization outlined above has any validity, there is no necessary reason to suppose that degrees of ambiguity do not exist or that subjects identifications might not more appropriately be recorded in terms of degrees of confidence. That recognition would be the crucial step in designing a new type of experiment to investigate the sentence comprehension processes, following the signal detection paradigm. Stimulus pairs should be drawn from sentence types possessing potential for ambiguity, such that for each ambiguous token there is a corresponding nonambiguous token. Use of semantic redundancy should allow this to be done with sentences involving both lexical and structural anomalies. Selected groups of subjects should then be instructed to respond by stating their confidence in detecting ambiguity on some scale (e.g., 1 to 5). Each group of subjects would be told to expect a different proportion of ambiguous to nonambiguous stimuli. Such an experiment could detect different response bias effects relative to lexical or structural types of information, and by comparing confidence ratings with response time might suggest the processes involved in distinguishing between ambiguity and nonambiguity.

A similar test designed to conform to the hierarchical categorization might make use of the names 'lexical', 'structural', and 'nonambiguous' as responses. If the cross categorization hypothesis is appropriate, then this change in form of response should not make it any easier for subjects to respond differentially in terms of both crossed factors simultaneously. Subjects would still respond dichotomously in terms of either potential or actual presence of ambiguity.

It seems appropriate to conclude with speculation, since the results of this research have raised more questions than they have answered. It has been noted that in this area of research, problems have arisen out of an uncritical acceptance of formal description and analysis. One could conclude that a more healthy scepticism is called for; however, that may not be carrying things far enough. If the hint that has been extracted from these results has a firm basis that the formal classification involved does not represent the categories in the manner in which subjects are aware of them, then the formal classification ought to be changed. The flow of information ought to be in two directions; where possible, formal distinctions ought to be given empirical verification, and where appropriate, formal distinctions ought to be derived from empirical results.

REFERENCES

Bever, T.G., Garrett, M.F. and Hurtig, R. The interaction of perceptual processes and ambiguous sentences. *Memory and Cognition* , 1973, *1*, 277-286.

Bever, T.G., Garrett, M.F. and Hurtig, R. Projection Mechanisms in Reading or When the Journal Review Process Fails. *Journal of Psycholinguistic Research*, 1976, *5*, 215-226.

Bradley, J.V. *Distribution-free Statistical Tests*. Englewood Cliffs, New Jersey: Prentice-Hall, 1968.

Broadbent, D.E. and Gregory, M. Human responses to classes of stimuli. *Nature*, 1962, *193*, 1315-1316.

Carey, P.W., Mehler, J. and Bever, T.G. Judging the veracity of ambiguous sentences. *Journal of Verbal Learning and Verbal Behaviour*, 1970, *9*, 243-254.

Carey, P.W., Mehler, J. and Bever, T.G. When do we compute all the interpretations of an ambiguous sentence ? In Flores d'Arcais and Levelt, 1970, 61-75.

Cattell, J. McK. *Psychological Research.* Lancaster, Pa.: Science Press, 1947.

Chomsky, N. *Aspects of the theory of syntax.* Cambridge, Mass.: The M.I.T. Press, 1965.

Chomsky, N. *Topics in the Theory of Generative Grammar.* The Hague: Mouton, 1966.

Christie, L.S. and Luce, R.D. Decision Structure and Time Relations in Simple Choice Behaviour. In R.D. Luce, R.R. Bush and E. Galanter (Eds.), *Readings in Mathematical Psychology Vol. 1.* New York; Wiley and Sons, 1963.

Cohen, J. and Cohen, P. Applied Multiple Regression/Correlation Analysis for the Behavioural Sciences. Hilsdale N.J.: Halsted Press, 1975.

Cooley, W. and Lohnes, P. Multivariate Data Analysis. New York: Wiley and Sons, 1971.

Dillon, P.J. Stimulus versus response decisions as determinants of the relative frequency effect in disjunctive reaction time performance. Journal of Experimental Psychology, 1966, 71, 321-330.

Draper, N.R. and Smith, H. Applied Regression Analysis New York: Wiley and Sons, 1966.

Flores d'Arcais, G.B. and Levelt, W.J.M. (Eds.) Advances in Psycholinguistics. Amsterdam: North-Holland Publishing Co., 1970.

Forrin, B. and Morin, R.E. Effects of contextual association upon selective reaction time in a numeral-naming task. Journal of Experimental Psychology, 1966, 71, 40-46.

Foss, D.J., Bever, T.G. and Silver, M. The comprehension and verification of ambiguous sentences. Perception and Psychophysics, 1968, 4, 304-306.

Garcia, E. Some Remarks on "Ambiguity" and "Perceptual Processes". Journal of Psycholinguistic Research, 1976, 5, 195-213.

Garrett, M.F., Does ambiguity complicate the processing of sentences?, In Flores d'Arcais and Levelt, 1970.

Gleason, H.A. Jr., Linguistics and English Grammar. New York: Holt, Rinehart and Winston, 1965.

Haggard, E.A. Intraclass Correlation with the Analysis of Variance . New York: Dryden Press, 1958.

Harman, G., Three Levels of Meaning. In D. Steinberg and L.A. Jakobovits (Eds.) *Semantics: An Interdisciplinary Reader.* London: Cambridge University Press, 1971.

Hick, W.E. On the rate of gain of information. *Quarterly Journal of Experimental Psychology,* 1952, 4, 11-26.

Kling, J.W. and Riggs, L.A. (Eds.) *Experimental Psychology.* New York: Holt, Rinehart and Winston, 1971.

Kornblum, S., Response competition and/or inhibition in two-choice reaction time. *Psychonomic Science,* 1965, 2, 55-56.

Lashley, K.S., The problem of serial order in behavior. In L.A. Jeffress (Ed.), *Cerebral Mechanisms in Behavior.* New York, Wiley and Sons, 1951.

Lefever, M.M. and Ehri, L.C., The relationship between field independence and sentence disambiguation ability. *Journal of Psycholinguistic Research,* 1976, 5, 99-106.

MacKay, D.G. To end ambiguous sentences. *Perception and Psychophysics,* 1966, 1, 426-436.

MacKay, D.G. Mental Diplopia: A model of speech perception. In Flores d'Arcais and Levelt, 1970.

MacKay, D.G. and Bever, T.G. In search of ambiguity. *Perception and Psychophysics,* 1967, 2, 193-200.

Mehler, J. and Carey, P. The interaction of veracity and syntax in the processing of sentences. *Perception and Psychophysics,* 1968, 3, 109-111.

Mistler-Lachman, J.L. Levels of comprehension in processing normal and ambiguous sentences. *Journal of Verbal Learning and Verbal Behavior,* 1972, 11, 614-623.

Morin, R.E., Konick, A., Troxell, N. and McPherson, S. Information and reaction time in 'naming' responses. Journal of Experimental Psychology, 1965, 70, 309-314.

Neimark, E.D. and Estes, W.K. (Eds.) Stimulus Sampling Theory. San Francisco: Holden-Day, 1967.

Overall, J.E. and Klett, C.J. Applied Multivariate Analysis. New York: McGraw-Hill, 1972.

Parsons, K.P. Ambiguity and the Truth Definition. Nous, 1973, 7, 379-394.

Patel, P.J. An Analysis of the Notion 'Structural Ambiguity'. Unpublished Doctoral Thesis, The University of Alberta, 1974.

Prideaux, G.D. Surface Structure Resolution of Structural Ambiguity, The University of Alberta, 1972.

Prideaux, G.D. and Baker, W.J. The Recognition of Ambiguity. Human Communication/Communication Humaine, 1976, 1, 51-58.

Quine, W.V.O., Word and Object. Cambridge, Mass. : The M.I.T. Press, 1960.

Quine, W.V.O. From a Logical Point of View. Cambridge, Mass.: Harvard University Press, 1964.

Siegel, S. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill, 1956.

Smith, E.E. Choice Reaction Time: An analysis of the major theoretical positions. Psychological Bulletin, 1968 69, 77-110.

Sokal, R.R. and Rohlf, F.J. Biometry. San Francisco: W.h. Freeman, 1969.

Stone, M. Models for choice-reaction time. Psychometrika, 1960, 25, 251-260.

Tatsuoka, M.M. Multivariate analysis : Techniques for educational and psychological research. New York: Wiley and Sons, 1971.

Van de Geer, J.P. Introduction to multivariate analysis for the social sciences. San Francisco: W.H. Freeman, 1971.

Welford, A.T. The measurement of sensory-motor performance: Survey and reappraisal of twelve years' progress. Ergonomics, 1960, 3, 189-229.

Winer, B.J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1971.

Woodworth, R.S. and Schlossberg, H. Experimental Psychology. New York: Holt, Rinehart and Winston, 1954.

Ziff, P. About What An Adequate Grammar Couldn't Do. Foundations of Language. 1965, 1, 5-13.

Ziff, P. Some Comments on Mr. Harman's Confabulations. Foundations of Language, 1967, 3, 403-408.

APPENDIX A

Subject's Instructions

This is an experiment to investigate how speakers of English analyze sentences to determine their meanings.

Most sentences are NON-AMBIGUOUS - they have only one meaning. For example;

Jerry is eating a ham sandwich.

Other sentences, however, may be AMBIGUOUS, that is they could be interpreted as having two or more possible meanings. For example;

The salesmen are going to hear a car talk today.

which might refer to a 'talking car' or a 'talk about cars'.

In this experiment, we are interested in how difficult it is to decide whether or not a sentence is ambiguous. In order to determine this, we will measure the amount of time it takes people to identify the ambiguous sentences in a list composed of both ambiguous and nonambiguous sentences. The list will be displayed on a video screen, one sentence at a time. Just before each sentence, there will be a warning display of X's. Your task will be to decide as quickly as possible whether or not each sentence is ambiguous.

When you have reached your decision, indicate this by responding YES or NO. When this is done the experimenter will press a button and the time taken to decide will be recorded by the computer. If you respond YES, you are asked to write a sentence corresponding to each of the interpretations that you saw, in the booklet provided. The sentence will remain on the screen for as long as you need, while you are writing the two interpretations.

If you are unable to reach a decision before one minute has passed, the sentence will be removed from the screen and we will go on to the next. So please respond as quickly as you canf bearing in mind that most of the sentences are ambiguous, but not all .

If you have any questions, please ask them now.

APPENDIX B

Stimulus Items

The following list contains all of the stimulus sentences used in both experimental procedures. In Experiment 1 the two possible clause orders of these sentences were included.

Lexical Amiguities

1. We got completely lost after taking the right turn at the corner.

2. He realized his blunder when he remembered that she couldn't bear children.

3. I was overly confident in thinking that the paper would cover everything.

4. She put on a clean apron before she carried out the orders.

5. When the colonel had finally gone, the soldiers put down their arms.

6. Because she was short of money, she borrowed from her closest friend.

7. As soon as the rain had stopped, the painter put on another coat.

8. The traffic cop liked his new position, although it involved some danger.

Surface Ambiguities

1. It didn't cause much embarrassment even when he laughed at the funeral.

2. Although I asked how old george was, I wasn't really interested.

3. She was having a rest while I was reading her mystery stories.

4. When he looked over the car, John noticed a large bloodstain.

5. The car kept stalling while they discussed their problems with the mechanic.

6. The stout doctor's wife stayed at home because the babysitter couldn't come.

7. Because of its importance, she sent the contract over a week ago.

8. Hoping for a big tip, I served the man with a smile.

Underlying Ambiguities

1. Although visiting relatives can be boring, I try hard to be patient.

2. I decided to go outside because I disliked smoking in the room.

3. He was exceeding his authority in ordering the police to stop drinking.

4. When they outlined their plans for us, we gave our complete approval.

5. Of all those told, the story of the salesman was the funniest.

6. Contrary to popular opinion, the French like opera as much as Italians.

7. Though he always left things in a mess, she liked his cooking.

8. Despite the surgery's success, the sight of the fireman was still awful.

Nonambiguous Sentences

1. He couldn't get a job after flunking out of school in the winter.

2. She remembered his name when she saw that he had a scar.

3. It didn't have the expected effect when they streaked during the ceremony.

4. They were robbing my house while I was working in the backyard.

5. I often miss my classes because I enjoy talking in the cafeteria.

6. She was exaggerating his importance in telling her friends about his job.

7. Despite the team's optimism, the injuries to the linemen were very serious .

8. He put on his new suit before he went out last night.

9. When the sun broke through, the game was in the final quarter.

10. When she looked at the thermometer, she saw that his fever had risen.

11. The fuses kept blowing while he was trying to fix the TV.

12. Of the few seen the birds at Slave Lake were the biggest.

13. When they examined the signature more closely, it was clearly a forgery.

14. He was always in debt to somebody because he couldn't save money.

15. The young couple like their new house although it had no garage.

16. Trusting to luck, I asked the police about my lost credit cards.

17. Because of its weight, he tried carrying the suitcase on his shoulder.

18. Though he appreciated the help they offered, he preferred doing it himself.

APPENDIX C

I. UNIVARIATE ANOVA'S ON TYPE USING SUBJECT STANDARD SCORES

A. LEXICAL AMBIGUITY

SOURCE	SS	DF	MS	F(MIXED)	F(FIXED)
MEAN	4.9492	1	4.9492	5.63 **	5.29 **
P	0.7014	1	0.7014	0.80	0.75
L	0.7542	2	0.7542	0.43	0.40
PL	2.4672	2	1.2336	1.40	1.32
S(PL)	31.6492	36	0.8791	0.94	0.94
R(PLS)	196.4540	210	0.9355		

B. SURFACE AMBIGUITY

SOURCE	SS	DF	MS	F(MIXED)	F(FIXED)
MEAN	0.0024	1	0.0024	0.0	0.0
P	0.0276	1	0.0276	0.04	0.03
L	5.7924	2	2.8962	4.29 **	2.69 *
PL	3.8389	2	1.9195	2.85 *	1.78
S(PL)	24.2775	36	0.6744	0.63	0.63
R(PLS)	226.1708	210	1.0770		

C. UNDERLYING AMBIGUITY

SOURCE	SS	DF	MS	F(MIXED)	F(FIXED)
MEAN	0.0108	1	0.0108	0.02	0.01
P	0.0020	1	0.0020	0.0	0.0
L	1.5732	2	0.7866	1.12	0.78
PL	0.6783	2	0.3391	0.48	0.34
S(PL)	25.3448	36	0.7040	0.70	0.70
R(PLS)	211.5656	210	1.0075		

D. NONAMBIGUOUS STIMULI

SOURCE	SS	DF	MS	F(MIXED)	F(FIXED)
MEAN	5.2005	1	5.2005	7.35 **	5.93 **
P	0.5119	1	0.5119	0.72	0.58
L	6.3600	2	3.1800	4.50 **	3.63 **
PL	7.9710	2	3.9855	5.64 **	4.55 **
S(PL)	25.4546	36	0.7071	0.81	0.81
R(PLS)	184.1423	210	0.8769		

+ 0.25 > P, * 0.10 > P, ** 0.05 > P, *** 0.01 > P

II. JOINT TESTS ON STIMULUS TYPES BY MANOVA

A. ALL STIMULUS TYPES

SOURCE	LN GENERALIZED VARIANCE	U-STATISTIC	APPROXIMATE F	DF
P	21.2543	0.9934	0.3446	4, 207
L	21.3225	0.9279	1.9737 **	8, 414
PL	21.3265	0.9242	2.0816 **	8, 414
S(PL)	21.7320	0.6161	0.7429	144, 827
R(PLS)	21.2476			

B. TYPES OF AMBIGUOUS STIMULI ONLY

SOURCE	LN GENERALIZED VARIANCE	U-STATISTIC	APPROXIMATE F	DF
P	16.0464	0.9962	0.2629	3, 208
L	16.0793	0.9640	1.2845	6, 416
PL	16.0736	0.9695	1.0811	6, 416
S(PL)	16.3895	0.7069	0.7094	108, 623
R(PLS)	16.0426			

III. DISCRIMINANT FUNCTION SIGNIFICANCE TESTS

FIGURE	FUNCTION	%VARIANCE	LAMBDA	CHI-SQUARE	DF	P
2	1	67.87	0.0079	55.68	40	0.05
	2	32.13	0.1248	25.93	19	0.199
3	1	93.44	0.3115	20.99	14	0.102
	2	6.56	0.8854	2.19	6	0.901
4	1	94.93	0.3690	18.44	12	0.103
	2	5.07	0.9255	1.43	5	0.921
5	1	71.05	0.3581	19.51	10	0.034
	2	28.95	0.7125	6.44	4	0.169
6	1	99.38	0.4271	17.44	4	0.002
	2	0.62	0.9918	0.17	1	0.681

IV. MULTIPLE REGRESSION

A. FULL MODEL

SOURCE	DF	SS	MS	F
REGRESSION	15	69.731	4.6487	11.998 ***
RESIDUAL	992	384.353	0.3875	

THE RESULTS FOR EACH VARIABLE WERE AS FOLLOWS:

VARIABLE	%VARIANCE	B	BETA	STD ERROR B	F
LKRT	0.631	0.07418	0.08382	0.02748	7.286 ***
SEX	1.165	-0.07305	-0.10884	0.01992	13.449 ***
BIAS	0.002	0.00029	0.00604	0.00214	0.018
ERROR	1.424	-0.10236	-0.15053	0.02525	16.436 ***
LAMB	0.522	0.02885	0.14953	0.01175	6.028 **
PARA	1.004	-0.10971	-0.11743	0.03222	11.594 ***
AMB	0.255	-0.12086	-0.15594	0.07038	2.949 *
TYPE	0.016	-0.06556	-0.08099	0.15417	0.181
SEQU	0.062	0.00287	0.02958	0.00339	0.717
STYPE	0.254	0.00619	0.10848	0.00362	2.931 *
SAMB	0.070	0.00306	0.05924	0.00339	0.812
ETYPE	0.079	0.02418	0.03118	0.02530	0.913
EAMB	3.571	0.15606	0.23213	0.02431	41.224 ***
BTYPE	0.001	-0.00016	-0.01381	0.00215	0.006
PTYPE	0.003	0.00605	0.00587	0.03278	0.034
CONSTANT		2.58286			

B. MAIN EFFECTS MODEL

VARIABLE	B	BETA	STD ERROR B	F
LKRT	0.05230	0.05910	0.02774	3.555 *
SEX	-0.06592	-0.09821	0.02030	10.545 ***
BIAS	0.00030	0.00633	0.00218	0.019
ERROR	-0.18493	-0.27195	0.02139	74.775 ***
PARA	-0.07919	-0.08476	0.03086	6.583 **
AMB	0.05406	0.06976	0.03552	2.317 +
TYPE	0.01118	0.01381	0.02580	0.188
SEQU	0.00080	0.00828	0.00292	0.076
CONSTANT	2.63643			

C. PARTITIONED ERROR BY STIMULUS TYPE INTERACTION

VARIABLE	B	BETA	STD ERROR B	F
S X ERROR	-0.22739	-0.23919	0.02811	65.422 ***
L X ERROR	-0.28994	-0.21390	0.04015	52.146 ***
N X ERROR	0.05025	0.03723	0.04074	1.521
SEX	-0.06630	-0.09878	0.01971	11.381 ***
PARA	-0.11938	-0.12779	0.02868	17.329 ***
CONSTANT	2.66176			

Mutual relations among the independent variables of the full regression model were evaluated in terms of the redundancy and suppression effects described in Cohen & Cohen (1975). These effects were identified firstly by noting whether the beta coefficient of any of the variables lay outside the theoretical bounds set by 0.0 and the simple correlation between that variable and the dependent variable; secondly, the thirteenth order partial among the independent variables were evaluated. In the present case, the following pairs of variables were found to be significantly related (at the .01 level): LKRT:SEX, LKRT:EAMB, SEQU:STYPE, AMB:LAMB, AMB:EAMB, TYPE:STYPE, LAMB:LKRT, LAMB:PARA, LAMB:TYPE, LAMB:PTYPE, and STYPE:SAMB. Following the procedures described in Cohen & Cohen (1975, p.460) the following pairs were identified as mutually enhancing variables: LKRT:EAMB, LAMB:LKRT, LAMB:PARA, STYPE:samb, and TYPE:STYPE. All other relationships were identified as exhibiting redundancy effects. In most cases, both the suppression and redundancy effects could be explained in terms of the relation by virtue of the coding of the respective variables. An exception is the relation between SAMB and STYPE, since it suggests that partialing on the distribution cf nonambiguous stimuli across trials is necessary before the sequence by TYPE discrimination interaction may be detected. This interpretation is supported by recollection that the lists were uniquely randomized for each experimental session.

TYPE			LEXICAL	SURFACE	UNDER-LYING	NONAMB-IGUOUS
POSITION 1	LIST 1	SUBS 1 . . 7	REPLICATES 1 - 6	REPLICATES 1 - 6	REPLICATES 1 - 6	REPLICATES 1 - 6
	LIST 2	SUBS 8 . . 14	...	...	...	...
	LIST 3	SUBS 15 . . 21	...	...	...	...
POSITION 2	LIST 1	SUBS 22 . . 28	...	...	...	...
	LIST 2	SUBS 29 . . 35	...	...	...	...
	LIST 3	SUBS 36 . . 42	...	...	...	...

Figure 1. - Data Structure

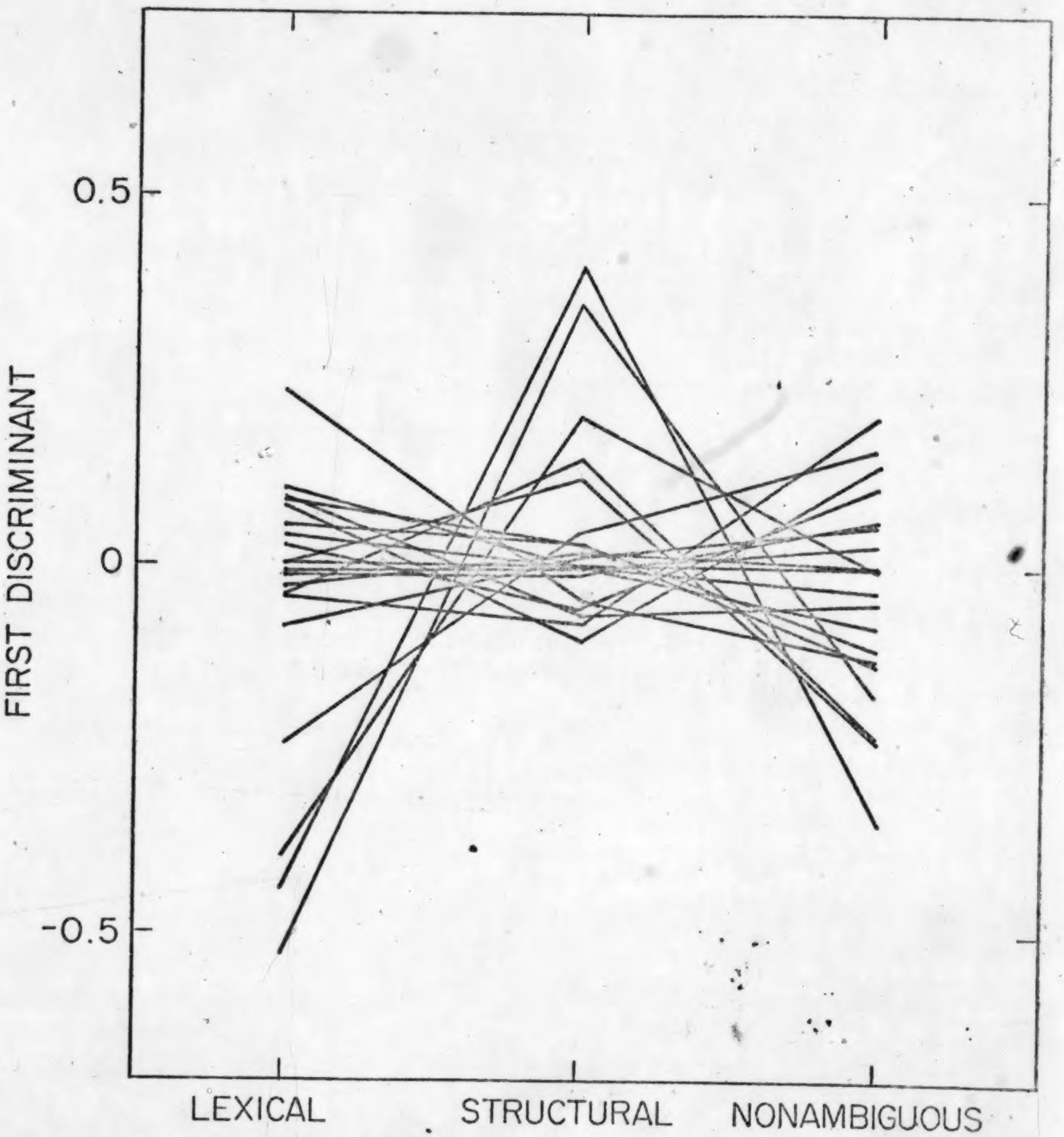

Figure 2. - Discriminators Response Profile

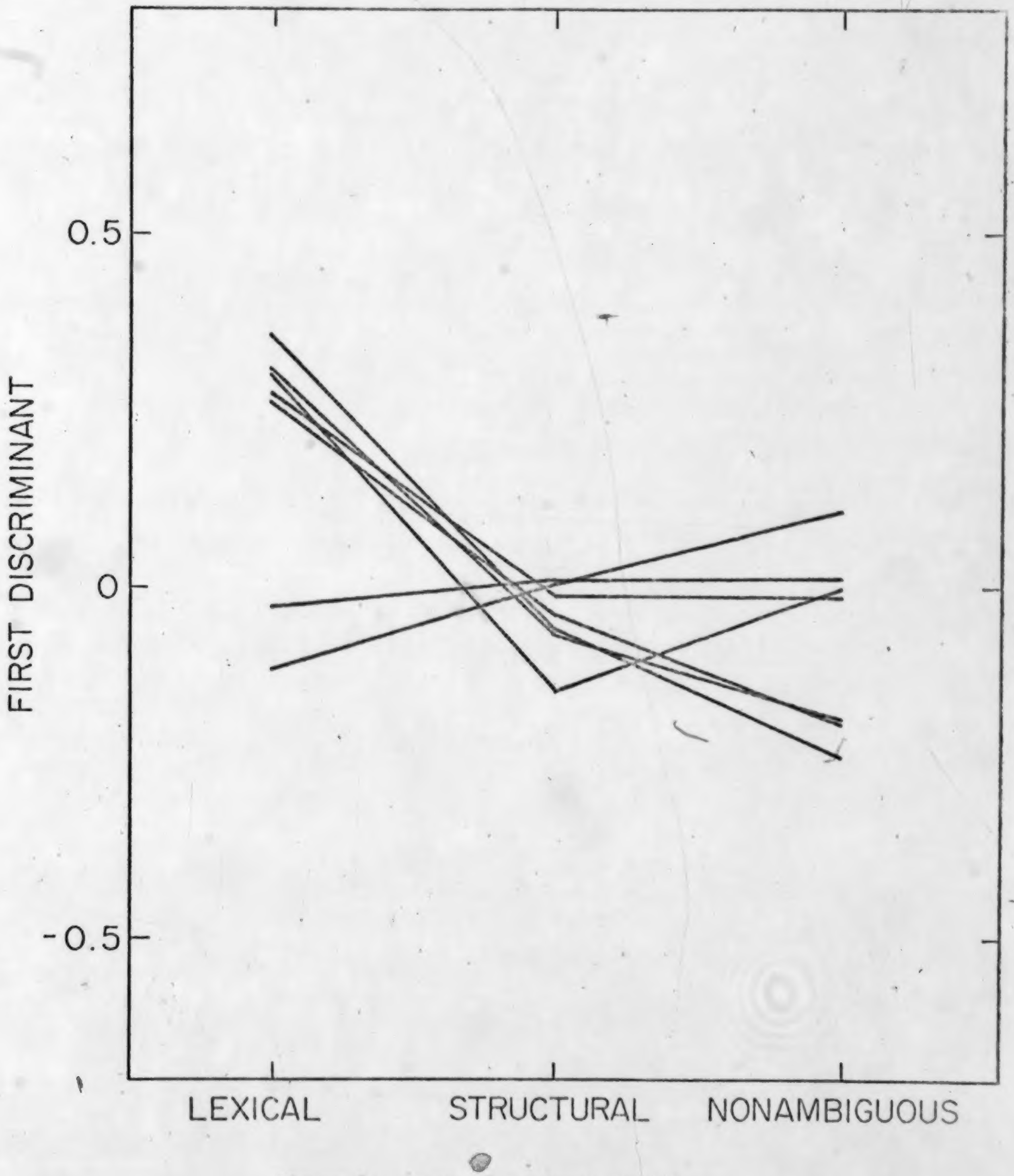

Figure 3. - Group 1 Response Profile

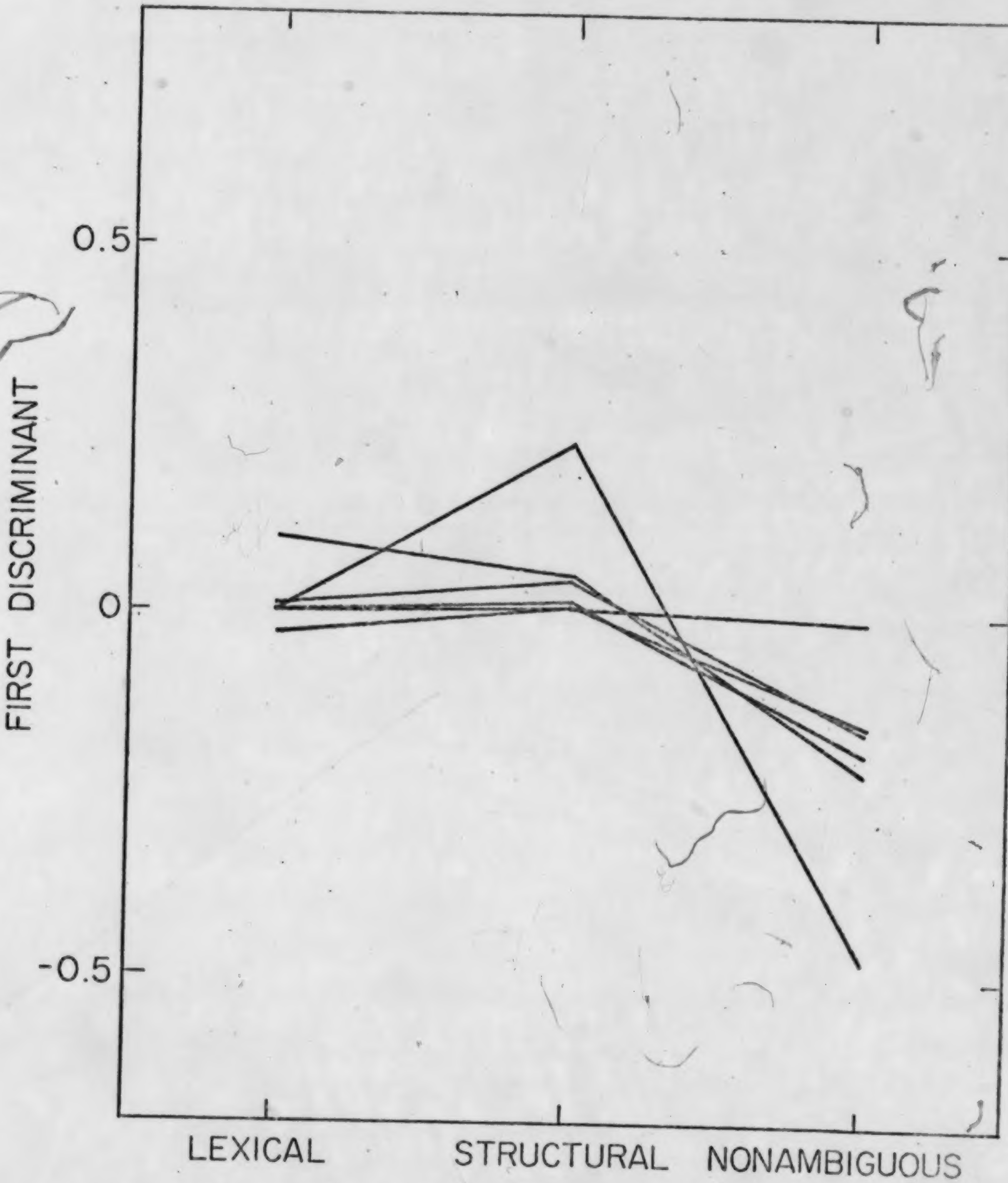

Figure 4. - Group 2 Response Profile

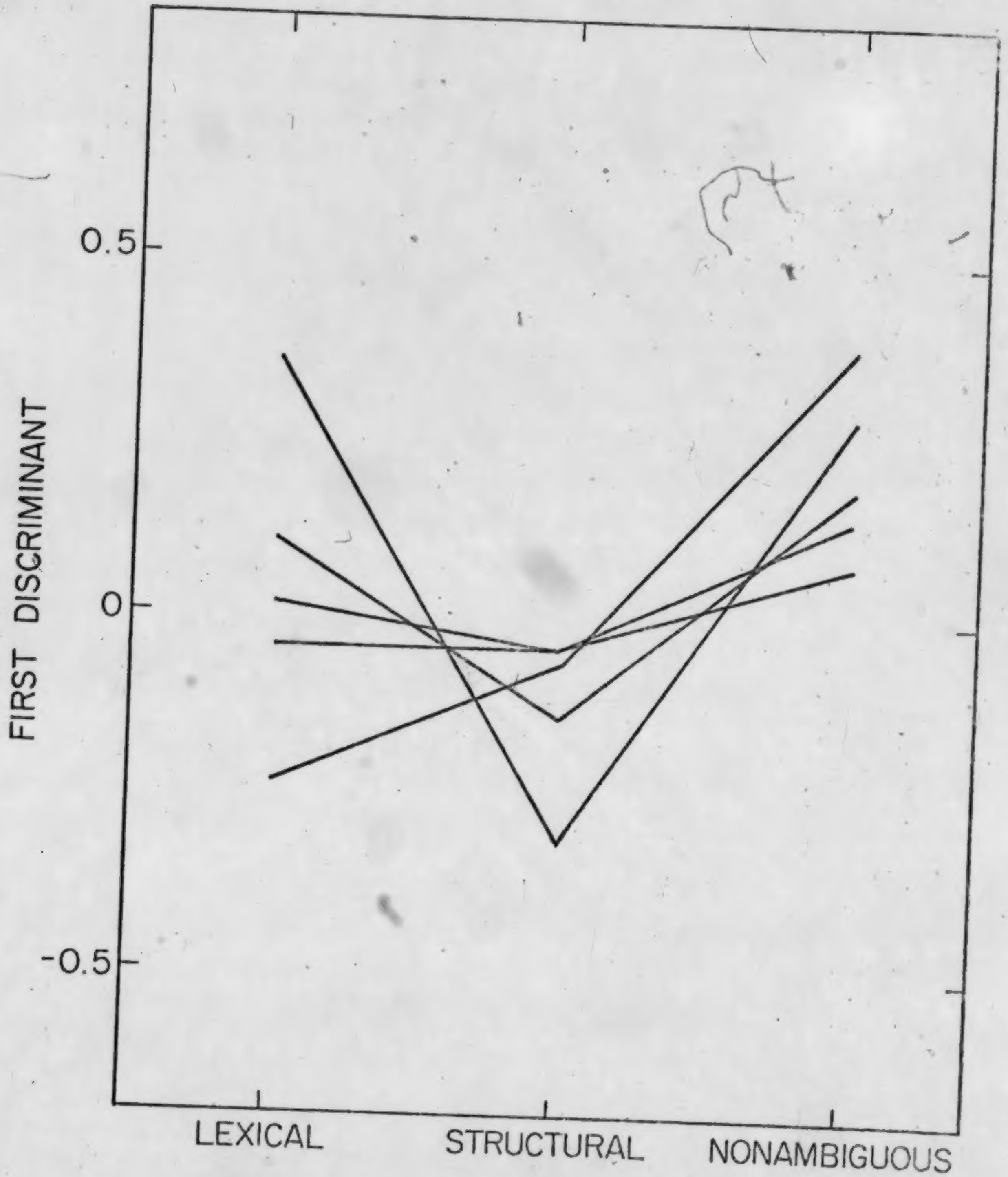

Figure 5. - Group 3 Response Profile

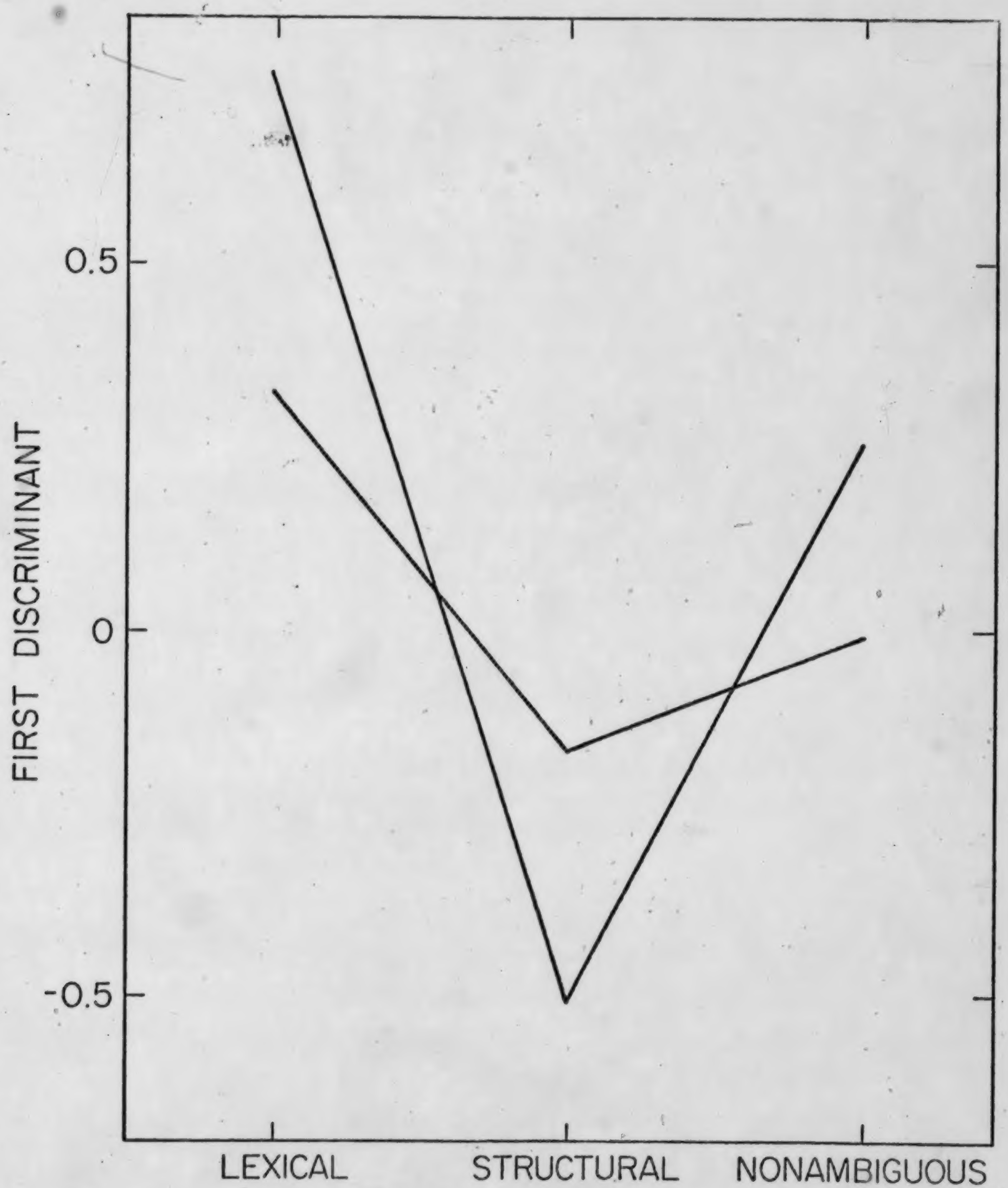

Figure 6. - Group 4 Response Profile

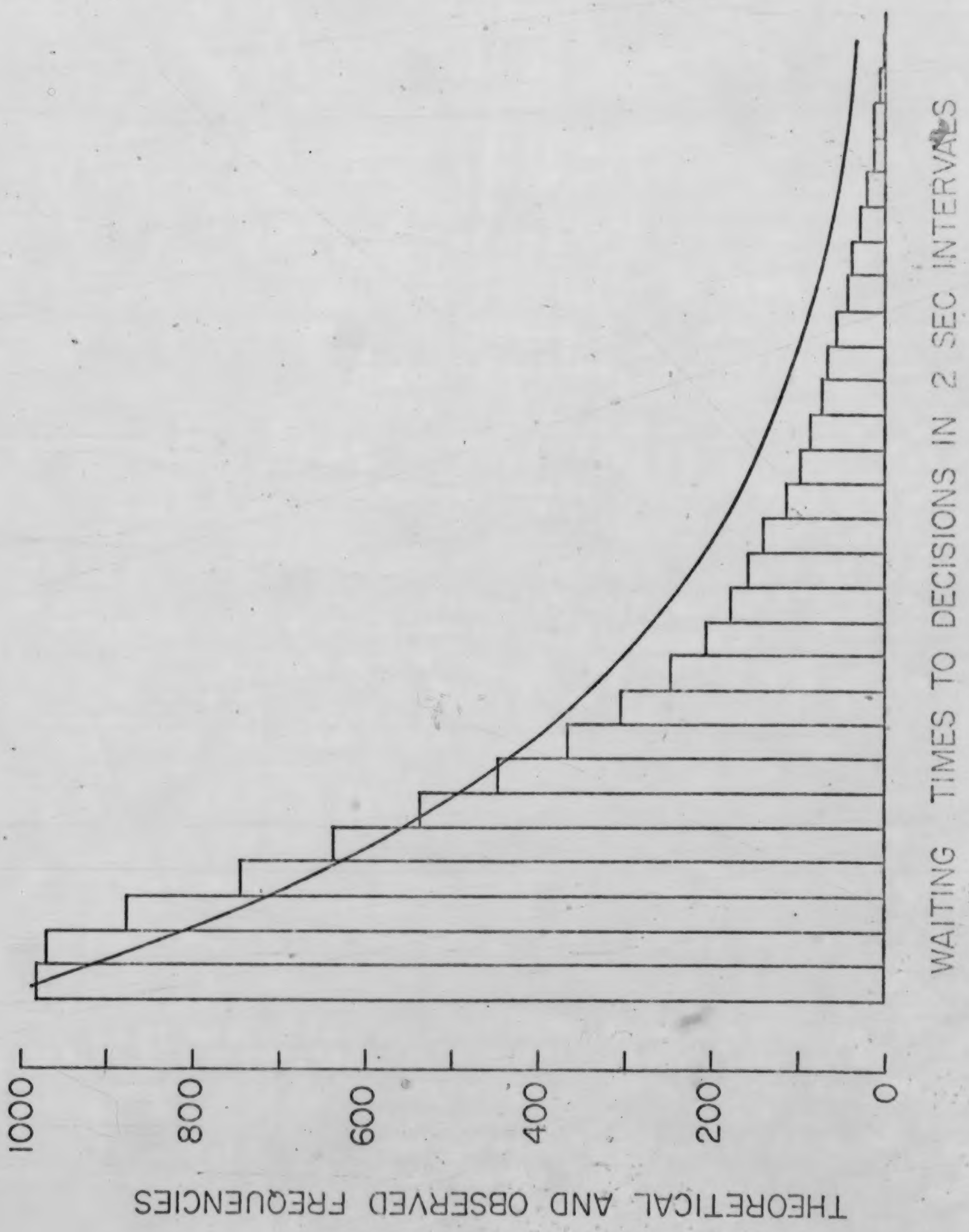

Figure 7. - Response Time Frequency Distribution